

# LIDAR DETECTION OF DEFOLIATION IN EASTERN HEMLOCK DUE TO HEMLOCK

## WOOLLY ADELGID

by

NICHOLAS KRUSKAMP

(Under the Direction of Marguerite Madden)

### ABSTRACT

LiDAR data are used to detect hemlock (*Tsuga canadensis*) defoliation due to hemlock woolly adelgid (*Adelges tsugae*) (HWA) in the Chattahoochee National Forest (CNF) located in Georgia. LiDAR data are used to quantify leaf area index (LAI) and fractional cover (fCover). Traditional field methods are used to validate LiDAR results with hemispherical photography of stands in varying stages of decline. Single linear regression results suggest that LAI ( $R^2 = 0.4355$ ) and fCover ( $R^2 = 0.4597$ ) are not well predicted from LiDAR. Multivariate principal component analysis using LiDAR variables and multivariate stepwise regression with ground data improve results for LAI ( $R^2 = 0.7307$ ) and fCover ( $R^2 = 0.4666$ ). Multivariate cluster analysis finds a significant relationship between ground and LiDAR derived clusters of high, medium, and low defoliation with 75% cluster agreement in classification of defoliation. This suggests that 3 health status ranks can be created from remote data.

INDEX WORDS: LiDAR, hemlock, remote sensing, forest canopy, leaf area index, fractional cover, defoliation, hemlock woolly adelgid

LIDAR DETECTION OF DEFOLIATION IN EASTERN HEMLOCK DUE TO HEMLOCK

WOOLLY ADELGID

by

NICHOLAS KRUSKAMP

BS Geography, University of Georgia, 2009

A Thesis Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of  
the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2011

© 2011

Nicholas Kruskamp

All Rights Reserved

LIDAR DETECTION OF DEFOLIATION IN EASTERN HEMLOCK DUE TO HEMLOCK

WOOLLY ADELGID

by

NICHOLAS KRUSKAMP

Major Professor:

Committee:

Marguerite Madden

Tomas Jordan

Jeff Hepinstall-Cymerman

Electronic Version Approved:

Maureen Grasso  
Dean of the Graduate School  
The University of Georgia  
August 2011

## ACKNOWLEDGEMENTS

I must first acknowledge my wife Kim for all the support and help she has given me – I could not have done this without her. I owe my deepest gratitude to my Major Professor, Marguerite Madden, and my committee, Tommy Jordan and Jeff Hepinstall-Cymerman. I thank my parents who always pushed me to be the best that I could be and to learn as much as possible. I am indebted to my large family of brothers, sisters, in-laws, nieces, nephews, cousins, aunts and uncles for the supportive and loving environment. Thank you to all those who work at CRMS – it was a place of learning, an academic support group, and a place to be safe. Thanks to Angela Mech, Brian Jackson, and Chris Strother who provided me with field guidance and data. Thanks to the Department of Geography, especially the staff and graduate students. I am fortunate to have some very close friends who have provided moral support and laughter, and I thank them for hanging in there through good times and bad. Lastly, thanks to my dog, Juno, for always reminding me to play.

The number of individuals who helped make this thesis possible amazes me, and my heartfelt thanks goes out to all those who have been involved in some capacity over the past two years.

## TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS .....	v
LIST OF TABLES .....	vii
LIST OF FIGURES .....	viii
CHAPTERS	
1 INTRODUCTION .....	1
Objectives .....	4
Study Area .....	5
Geospatial Data .....	11
Scientific Significance .....	13
2 LITERATURE REVIEW .....	15
Eastern Hemlock .....	15
Hemlock Woolly Adelgid .....	19
Effects of Hemlock Decline .....	21
Ground Measurements of Leaf Area Index .....	22
LiDAR .....	24
LiDAR Applications in Forest Studies .....	27
3 METHODOLOGY .....	31
Field Data Collection .....	31
Field Data Processing .....	37
LiDAR Data Processing .....	45

Statistical Analysis Methods.....	50
4 RESULTS .....	56
Linear Regression Results.....	56
PCA and Stepwise Multivariate Regression Results .....	65
Multivariate Health Status Clustering Results .....	70
5 DISCUSSION.....	76
Environmental Considerations.....	76
Data Equation Considerations.....	77
Canopy Geometry in ALS Measurements .....	79
Data Trap Size and Sampling Rate .....	79
Prediction Power of Additional ALS Variables.....	81
Agreement among ALS, Ground and Field HSr.....	82
6 CONCLUSION.....	83
REFERENCES .....	87

## LIST OF TABLES

	Page
Table 1: Geospatial Data.....	12
Table 2: HPE, VPE, and PDOP of GPS Points .....	34
Table 3: Field Survey Data .....	37
Table 4: Can-Eye Software Settings .....	41
Table 5: RMSE and Z Bias of QTM DEM Surfaces .....	49
Table 6: List of All Ground and LiDAR Variables .....	51
Table 7: R <sup>2</sup> Values for ALS and Ground LAI .....	57
Table 8: LAI Single Linear Regression Results .....	59
Table 9: R <sup>2</sup> Values for ALS and Ground fCover .....	61
Table 10: fCover Outlier Analysis Variables.....	62
Table 11: fCover Single Linear Regression Results .....	64
Table 12: PCA Eigenvalues and Proportions.....	66
Table 13: PCA Factor Loading Matrix .....	67
Table 14: LAI Multivariate Stepwise Regression Results .....	68
Table 15: fCover Multivariable Stepwise Regression Results .....	69
Table 16: Field and GS HSr Contingency Table .....	71
Table 17: Field and ALS PC HSr Contingency Table .....	73
Table 18: GS and ALS PC HSr Contingency Table .....	74
Table 19: GS and ALS PC HSr for HCA-IDs .....	75

## LIST OF FIGURES

	Page
Figure 1: Chattahoochee Forest Age Distribution .....	6
Figure 2: The Chattahoochee National Forest .....	9
Figure 3: Four Main Study Areas .....	9
Figure 4: HCAs and Sites .....	10
Figure 5: Eastern Hemlock Native Range .....	16
Figure 6: Healthy Eastern Hemlock Branch .....	18
Figure 7: Counties Reporting Infestation, 2010.....	20
Figure 8: Infested Eastern Hemlock Branch.....	21
Figure 9: Airborne Laser Scanner System .....	26
Figure 10: Full Waveform and Discrete LiDAR .....	26
Figure 11: Camera Calibration.....	39
Figure 12: Extracting X Y Pixel Values for Calibration.....	40
Figure 13 : Can-Eye Classified Images .....	42
Figure 14: 20-m Data Traps on Digital Elevation Model Surface.....	47
Figure 15: R <sup>2</sup> Values for ALS and Ground LAI.....	58
Figure 16: ALS and Ground LAI Scatter Plot and Linear Fit.....	59
Figure 17: R <sup>2</sup> Values for ALS and Ground fCover.....	61
Figure 18: ALS and Ground fCover Scatter Plot and Linear Fit .....	64

## CHAPTER 1

### INTRODUCTION

Airborne discrete return LiDAR (Light Detection and Ranging) is quickly becoming a new standard remote sensing technique used in visualizing and analyzing forest structure. Remote measurement and characterization of forest canopy and structure in moderate to high biomass forest, such as the coniferous eastern hemlock (*Tsuga canadensis*) and Carolina hemlock (*Tsuga caroliniana*), have proven difficult (Lefsky et al. 1999, Watson 1947). Field methods for collecting data are often impractical to cover broad spatial scales, while traditional optical remote sensing techniques can cover larger areas, but may lack the detail needed for fine scale analysis and often cannot see beneath forest canopies (Chen and Cihlar 1996). LiDAR is a remote sensing technology that uses laser pulses to determine the range of a target from the sensors. The ability of LiDAR to have multiple returns from a single laser pulse provides a way to look at not only the top of forest canopies, as with traditional photogrammetric techniques, but also to see through to subcanopy and forest floor. This additional dimension, a vertical distribution of data in conjunction with the traditional horizontal distribution, allows researchers to view and understand forest structure in 3D (Lim et al. 2003). LiDAR has been applied in a number of previous studies to estimate forest metrics (Brandtberg et al. 2003, Lefsky et al. 1999, Means et al. 1999, Lefsky et al. 2002, Kane et al. 2010). The most common LiDAR used for forest studies has been large footprint full waveform LiDAR, first and last discrete return LiDAR, and more recently, high density multiple return (up to 5 returns per pulse) discrete LiDAR.

Hemlock woolly adelgid (*Adelges tsugae*) (HWA), a small exotic and invasive insect thought to have been introduced from Asia via infested lumber, has been the cause of hemlock decline along the east coast of the United States for over five decades (Orwig, Foster and Mausel 2002). HWA targets all species of hemlock, but eastern and Carolina hemlock have no environmental or genetic defenses against the insect. HWA feeds at the base of needles on the tree, and overwhelms the tree's natural systems by interrupting resource flow to and from needles and through infestation of hundreds of thousands of insects on each tree. The infestation causes severe needle loss, and if left untreated, can lead to tree mortality within ten to fifteen years. Eastern hemlock creates a cool, moist forest ecosystem with its dense needle and limb structure that extends from the top of the canopy to the ground (Brisbin 1970). It is only found along the eastern coast of North America, from Canada to northern Georgia. This species typically grows in shaded valleys in riparian zones, and contributes to the high biodiversity found in the southeast of the United States. It is essential to understand changes in the canopy structure of these trees related to foliage density and solar radiation penetration to understand how the species responds to disturbances such as HWA at multiple scales—from the individual hemlock tree to the ecosystem level (Coops et al. 2007). Comparing LiDAR data among stands of varying health may provide information to quantify the ecological effects of HWA and the efficacy of HWA control measures and treatments, provide a temporal comparison in future studies, and broaden understanding of forest ecosystem responses to exotic pest infestations.

Leaf Area Index (LAI) is an integral variable in ecosystem modeling that estimates the area of foliage cover over an area of ground. LAI is generally measured from 0 – 6, with 0 being representing bare ground and 6 a dense forest (Jonckheere et al. 2004). LAI is sensitive to the response of, and provides an index with which to compare, forest stands change due to

disturbance. Destructive measures of LAI require either the collection of leaves from an area with collection nets or destructive removal of foliage from trees, and scanning all leaves to calculate the total leaf area. Indirect measurements can be employed with similar accuracy in order obtain effective LAI measurements (Riaño et al. 2004). There are a number of different ways to estimate LAI from the ground, including specialized LAI instruments (e.g., LICOR LAI2000) which measures available light under canopy, or hemispherical photography which optically assesses canopy cover relative to open sky. The latter method requires only a high quality camera equipped with a fish-eye lens, and software able to analyze vegetation and sky areas of images.

Fractional Cover (fCover) is a variable that quantifies the amount of ground covered by vegetation in a given area. It represents the open canopy percentage and exposed ground as a means of quantifying the amount of total vegetative cover, or lack thereof, in forest canopies. From the ground, fCover is expressed as a percentage of canopy covering an area, or the openness of the canopy (Hopkinson and Chasmer 2009). From LiDAR, it can be expressed as a fraction of canopy laser returns to ground laser returns (Morsdorf et al. 2004).

While field based methods for measurement of LAI and fCover are effective, they are only effective at local scales due to the high cost of manual labor and the need for suitable ground conditions and access. LiDAR LAI and fCover measurement, in comparison, allows for analysis on broad landscape and regional scales (Lefsky et al. 2002). Correlation between field-based and LiDAR-based measurements may therefore provide an effective model which can accurately assess these metrics over multiple scales, while helping to identify varying degrees of infestation in hemlock forest stands.

## Objectives

In this study, the results of regression analysis between ground based and LiDAR based forest stand characteristics offer further understanding about the changes in forest canopy due to the invasion of hemlock woolly adelgid. These analysis is performed from a sample of hemlock tree stands within the biologically diverse Chattahoochee National Forest, Georgia, United States. The primary objective of this research is to identify relationships between ground and LiDAR based metrics for characterizing the structure of hemlock stands affected by HWA infestation, and to use those relationships to distinguish between stands of varying intensity of defoliation due to infestation. Results produced here focus on establishing statistical relationships between laser derived data and ground based measurements, and add to a growing list of research on the topic of LiDAR study of forest canopy (Lim et al. 2003). In building this statistical relationship, attempts are made to determine distinct groups, or clusters, of LiDAR values representing the intensity of HWA infestation in a given stand—values that represent healthy and unhealthy stands of trees. The methods from this study can be applied to determine the efficacy of treatment, for comparison of temporal changes over time, and broaden the scope of study from a local scale to a regional scale.

The specific objectives of the research are:

1. To analyze LiDAR data of HWA-infested stands in the Chattahoochee National Forest of north Georgia. Approximately 30 sites with damage that range from little noticeable defoliation, to stands that have severe damage and defoliation and high mortality are assessed. LiDAR based metrics derived are LAI, fCover, mean canopy height, height percentiles, average slope, average elevation, in canopy vegetation fraction (ICVF), and descriptive LiDAR point return statistics.

2. To collect field based data on selected study sites which coincide with LiDAR data that are representative of hemlock stands along a spectrum of infestation stages. Sites are assigned a health status ranking (HSr) from 1 (healthy) to 3 (very unhealthy). LAI, fCover, average stem density per site, species inventory, and presence of HWA are collected for each site.
3. To use the collected data to create empirically based single linear regression models which explain ground measurements of LAI and fCover using LiDAR derived measurements, and to use principal components analysis and stepwise regression to perform multivariable regression and test if the inclusion of additional LiDAR variables improves the results of regression.
4. To calculate HSr rankings through Kmeans cluster analysis of LiDAR and ground data, and test if there are statistically significant relationships between those clusters and the HSr assigned from field work.

### Study Area

The study area for this research is in the Chattahoochee National Forest (CNF) of north Georgia within the southern end of the Appalachian Mountains. Elevations generally range from 250 to over 1200 m and the forest includes the state's highest peak, Brasstown Bald, at 1458 m. The CNF covers approximately 3200 square kilometers and eighteen counties and is comprised of a variety of forest age classes (Figure 1) (CNF 2011).

Named after the Chattahoochee River, a major river whose headwaters originate in the Georgia Mountains, the CNF began as a purchase of approximately 12 square kilometers in 1911 by the USFS. Prior to this land purchase, the area underwent heavy logging, often consisting of

clear cutting large tracts of land, which is primarily responsible for the relatively young ages of trees found there today.

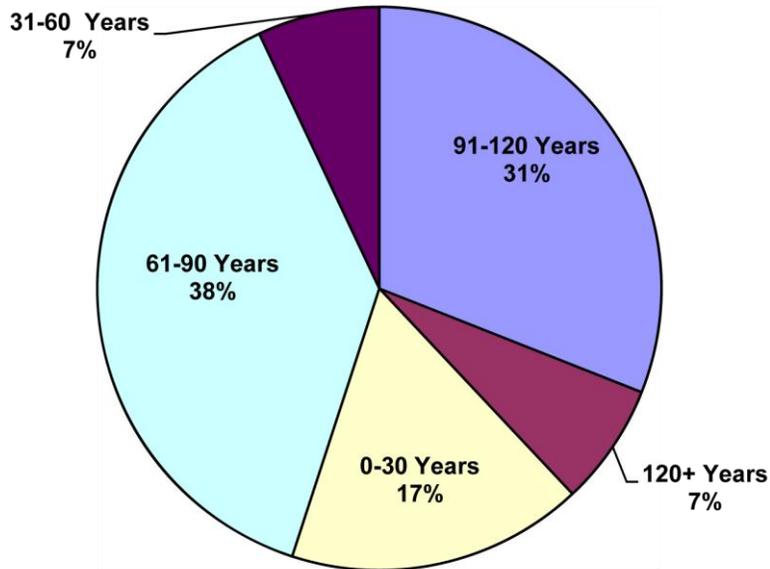


Figure 1: Age distribution of forests in Chattahoochee National Forest (CNF 2011).

After 1911, additional land purchases were negotiated and the trees on national forest land were allowed to regrow. Logging continued in the forest, but under much tighter restrictions towards stewardship of the land (Davis 2000). In 1959, the CNF was merged with a central Georgia land purchase, the Oconee National Forest, and the two became collectively known as the Chattahoochee-Oconee National Forest. However, hemlocks only grow on the CNF portion of the National Forest. Today, the CNF encompasses ten wilderness areas as well as a Wild and Scenic River, The Chattooga River. The forest is split into four ranger’s districts: Blue Ridge, Chattooga River, Conasauga, and Oconee Ranger Districts with forest headquarters in Gainesville, Georgia.

In response to the growing need for research on the effects of HWA, as well as the need to treat HWA effectively, the USFS has designated certain areas of hemlock forest within the CNF as Hemlock Conservation Areas (HCA). These areas have been deemed either biologically or recreationally important, and often are sites for both entomological and silvicultural research. In many cases, the areas are sites of collaborative efforts from many different universities and research institutions to test both chemical and biological treatment of HWA, and maintain certain areas as controls.

With assistance from University of Georgia researchers, four HCAs were selected for study: HCA #72 Slaughter Creek; HCA #71 Dick's Creek; HCA #145 Soque River; and HCA #29 Panther Creek (Figure 4). Each area was visited and assessed for appropriateness of study and availability of LiDAR data. Each site was assigned a unique identifier as it was surveyed (HCA-ID), a value that corresponds to the HCA it is in and the order it was surveyed (e.g., HCA ID 29-01 is HCA #29 site 01). These sites represent a total area approximately 50 km wide east to west and 10 km north to south. Study sites generally follow rivers and streams in the HCA. In many cases, sites were located in flat floodplains and on slopes bordering the streams. Sites were hemlock dominated stands with mixed pine and a limited number of mixed deciduous trees and were generally devoid of understory or low lying vegetation. Study areas were located away from urban development or other human influences, and consist of generally dense forest canopy (Figure 4). The forest cover, however, is not homogeneous, and often changes composition quickly as one departs upslope from water. Hemlock habitat in Georgia is limited to low lying, shady areas, mainly along bodies of water – as exemplified in the selected study sites.

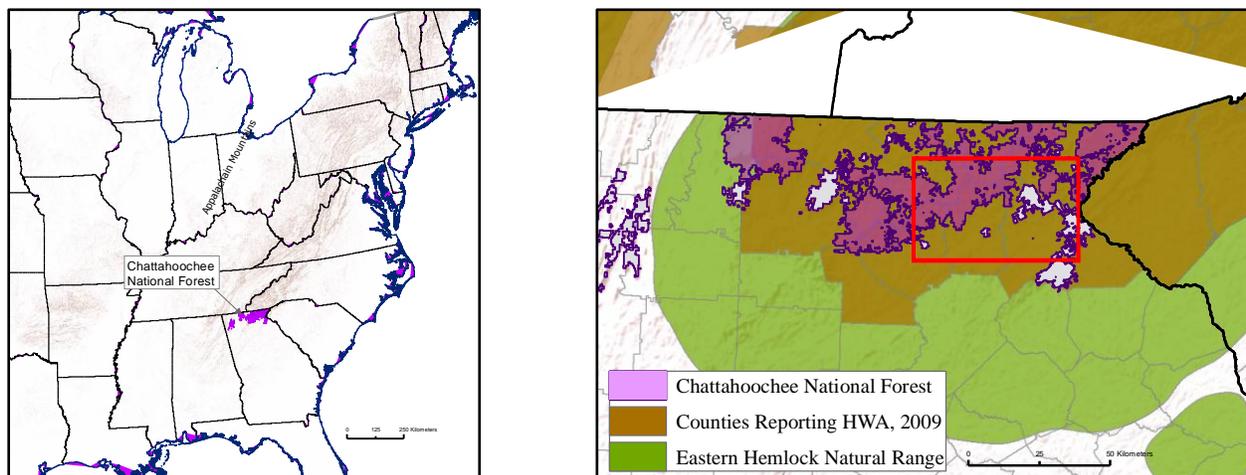


Figure 2: The Chattahoochee National Forest is located on the southern end of the Appalachian Mountains. The red rectangle shows the extent of the study sites used in the study (Figure 3).

The north Georgia Mountains present an appropriate location for this study for a number of reasons. North Georgia represents the southern extent of eastern hemlocks' natural range, as well as the range of Carolina hemlock. These sites lay along the southern end of the known HWA infestation in Georgia (Figure 2). In 2002, Rabun County, Georgia's northeastern most county, became the first in Georgia to report HWA infestation. The earliest report of HWA from a county within the range of the study areas is 2005, with one county reporting as late as 2007. Since then, the infestation has spread south and west, with a total of 11 counties reporting HWA as of 2009 (USFS 2011). While the counties that comprise the CNF all report infestation at this time, due to the sporadic nature of HWA spread, there is a mix of both healthy and unhealthy hemlock stands found in the CNF. In contrast, resource managers at the Great Smoky Mountains National Park (GRSM) observe that almost 100% of their hemlock stands are severely infected, and are experiencing high mortality rates (Tom Remaley, GRSM Ecologist, personal correspondence). The selected study area of north Georgia, therefore, is better suited for the field component of this study, and provides a range of trees in various stages of decline. In addition,

due to lower elevations, southern latitudes, and a milder climate, hemlock habitat in north Georgia is generally accessible year round.

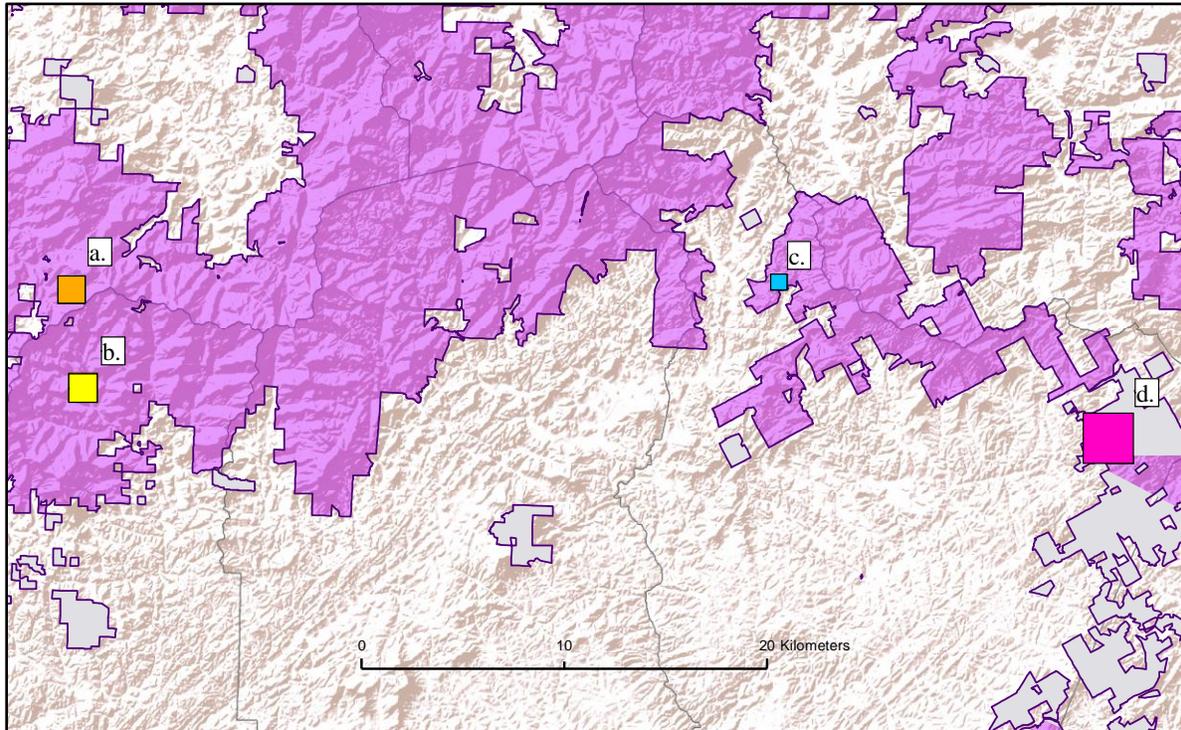


Figure 3: Four Hemlock Conservation Areas were selected for this study: (a) Slaughter Creek, HCA # 72, (b) Dick's Creek, HCA # 71, (c) Soque River, HCA # 145, and (d) Panther Creek, HCA # 29. Each area can be seen in detail in Figure 4.

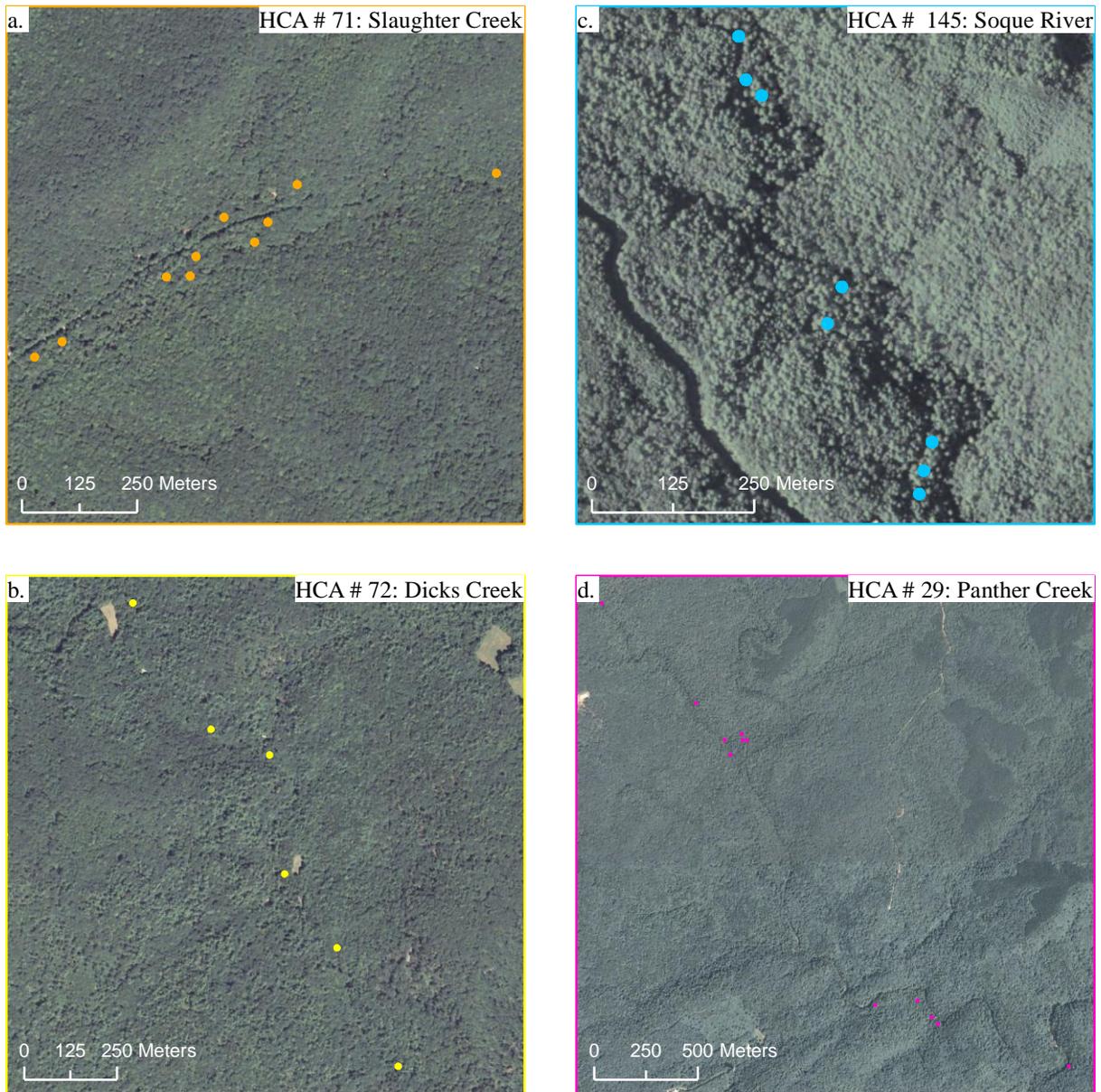


Figure 4: All sites are shown superimposed over U.S. Department of Agriculture National Agricultural Imagery Program (NAIP) 2010 imagery. Each circular polygon represents a 20-m diameter study site. There are 10 sites for HCA # 71, 6 sites for HCA # 72, 8 sites for HCA # 145, and 12 sites for HCA # 29 as indicated by the points in each photo.

## Geospatial Data

The available LiDAR data were flown by the vendor Photo Science, Incorporated (PSI), based in Lexington, Kentucky. The data were collected using a Leica Geosystems ALS50 airborne laser scanner (ALS) and an Optec ALTM Gemini airborne laser scanner in leaf-off conditions in spring 2010. This Leica system has a 150 kHz maximum pulse rate and a four return range detection system, with a laser frequency of 1064 nm in the near-infrared spectral region. The Optec system has a 167 kHz maximum pulse rate and a four return range detection system, also with a laser frequency of 1064 nm. Collected as part of a U.S. Geological Survey (USGS) grant for Gainesville State College, the data meets the USGS V12 LiDAR specifications. The nominal pulse spacing is 1 m, though this varies in the along-track and across-track directions (with a wider spacing between sets of scan lines). Data were collected in snow free conditions with water bodies at or below their normal levels. Vertical accuracy is  $\pm 15$  cm and horizontal accuracy is less than 1 m. The data were delivered in classified raw .LAS files tiles measuring 1.5-km by 1.5-km with four vendor provided classes: 1 – Unclassified, 2 – Bare Earth, 7 – Noise, 9 – Water, and 12 – Overlap. The raw .LAS tiles contained an average of over 6 million points per tile. These tiles were provided for this study by Gainesville State College in conjunction with PSI and the USGS. A total of 7 tiles were received and processed according to the methods described below using QCoherent LP360 software.

Four band ortho imagery also were collected in tandem with the LiDAR data as part of the USGS grant. The imagery was collect at approximately 15-cm ground pixel resolution and the four bands were red, green, blue, and near infrared. These images can be viewed in false color infrared for clearer distinction between evergreen and deciduous vegetation.

The U.S. Department of Agriculture (USGS) National Agricultural Imaging Program (NAIP) 2010 imagery for the state of Georgia was collected in leaf-on conditions in the summer of 2010. These data provided 3-band true color imagery with a ground pixel resolution of approximately 1 m. This ortho rectified dataset is used as a base layer and for comparisons between the four band leaf-off imagery.

Table 1: A list of geospatial data and sources used in this study.

<b>Data</b>	<b>Resource, Agency:</b>
LiDAR .LAS tiles (7 Tiles)	Photo Science, Inc., Gainesville State College, USGS
<b>Imagery:</b>	
4 band ortho image tiles	Photo Science, Inc., Gainesville State College, USGS
NAIP 2010	USDA FSA APFO ESRI ArcServer Database
<b>Land Cover:</b>	
Land cover and forest type classification	Brian Jackson, USFS CNF Headquarters
HCA database	Brian Jackson, USFS CNF Headquarters
<b>Site Level Information:</b>	
HCA treatment database and maps, treatment and control sites	Angela Mech, University of Georgia Warnell School of Forestry and Natural Resources
Field work database	Personally collected data

Several GIS databases of both vegetation cover classification and land cover classification were provided by USFS CNF Headquarters, Gainesville, Georgia. This database contained over 70 land cover classes, approximately 100 forest type codes, as well as over- and understory species codes. These layers contain both estimated stand ages as well as detailed primary/secondary/tertiary dominant vegetation classification. For this study, forest type classes of 5 – Hemlock, 8 – Hemlock-Hardwood and 41 – Cove Hardwood-White Pine-Hemlock or two species codes TSCA and TSCA2 (eastern hemlock and Carolina hemlock, respectively) were

considered to be valid potential study site land cover and forest type classes. CNF Headquarters also provided the latest GIS polygon layers of all HCAs within the national forest boundary. Where applicable, this layer contained information on types of treatment (chemical, biological, or both) being used in a given HCA.

Angela Mech, a University of Georgia Warnell School of Forestry and Natural Resources graduate student involved in HWA research and biological control, provided four maps containing treatment and control points for the HCAs used in this study. These maps were imported into ESRI ArcMap and georectified. Points were extracted from these maps for use as guides in the field when identifying appropriate study sites.

Field data were collected in this study and compiled into a research database. This database consisted of a written site assessment, a count of all main canopy stems within a 10-m radius of the center of a study site, hemlock species (only) stem count within 10-m, and all relevant GPS information needed to reference the site. Each site was qualitatively assessed to determine a hemlock health status ranking (HSr). This assessment was based on the presence of HWA at a site, the amount of defoliation, and the height above ground defoliation.

#### Scientific Significance

This study contributes to the growing need for information on the spread of HWA throughout the eastern coast of the United States. While optical remote sensing techniques have been widely used to assess damage and identify areas of infestation, this is one of the first studies to incorporate the application of LiDAR to examine the difference in metrics between stands of varying stages of infestation. Understanding canopy changes, and being able to visualize them vertically, can provide a better understanding of the progression of the disturbance, and a way to view loss of habitat in the canopy. This study also demonstrates the vertical distributions, and

differences therein, of canopy and shows how that distribution changes as infestation progresses. While LiDAR alone may not be able to predict areas of infestation with certainty due to the high variance of canopy form between individual trees and variable microhabitats of specific locations, it may be used to aid in the identification and visualization of hemlock sites undergoing defoliation. Finally, while other studies add to the growing body of research in the applications of LiDAR for forest metrics by looking at differences in sites due to age or species type, this study is unique in that it looks at differences in stands due to a disturbance, and compares sites based on the progression of that disturbance.

## CHAPTER 2

### LITERATURE REVIEW

#### Eastern Hemlock

The eastern hemlock (*Tsuga canadensis*), a member of the pine family, is one of four known hemlock species that grows in North America. In total, there are only ten hemlock species known in the world (Brisbin 1970). Two species of hemlock, the western hemlock (*Tsuga heterophylla*) and mountain hemlock (*Tsuga mertensiana*), grow in the western and north-western areas of the United States and Canada, while the Carolina hemlock (*Tsuga caroliniana*) is found exclusively in the southeastern part of the United States. The eastern hemlock naturally ranges from Nova Scotia, Canada in the north, southward along the coast and Appalachian Mountains to Georgia and Alabama, U.S. to the south. It is also found in the Lake States as far west as Michigan and Wisconsin (Figure 5) (Burns and Honkala 1990, Brisbin 1970). Though there exists a mix of eastern hemlock and Carolina hemlock in the study area, the follow discussion is about eastern hemlock only, since it is believed that it constitutes the majority of hemlock in Georgia.

Eastern hemlocks are unique in that they are one of the most shade tolerant tree species known, with seedlings being able to survive with less than 5 percent of full sunlight (Carey 1993). Typically, this species grows slowly throughout sapling and pole stages, and is able to withstand frequent suppression and release cycles during growth. Generally, hemlock germination rarely occurs without heavy overstory, due to moisture stress caused by germination in the open (Burns and Honkala 1990). It is not uncommon to find individual trees over 100 years old with a diameter at breast height (DBH) less than 5 cm. Mature eastern hemlock DBH

and height varies spatially, but generally diameters range between 89 and 102 cm and heights are greater than 30 m. Typically, DBH and height is slightly greater in the Southern Appalachians than in New England or the Lake States (Burns and Honkala 1990). While in sapling and pole stages, hemlocks tend to have a dense conic or pyramidal crown with the lower branches very close to the ground; while in mature stages they tend to have ragged crowns (Brisbin 1970).

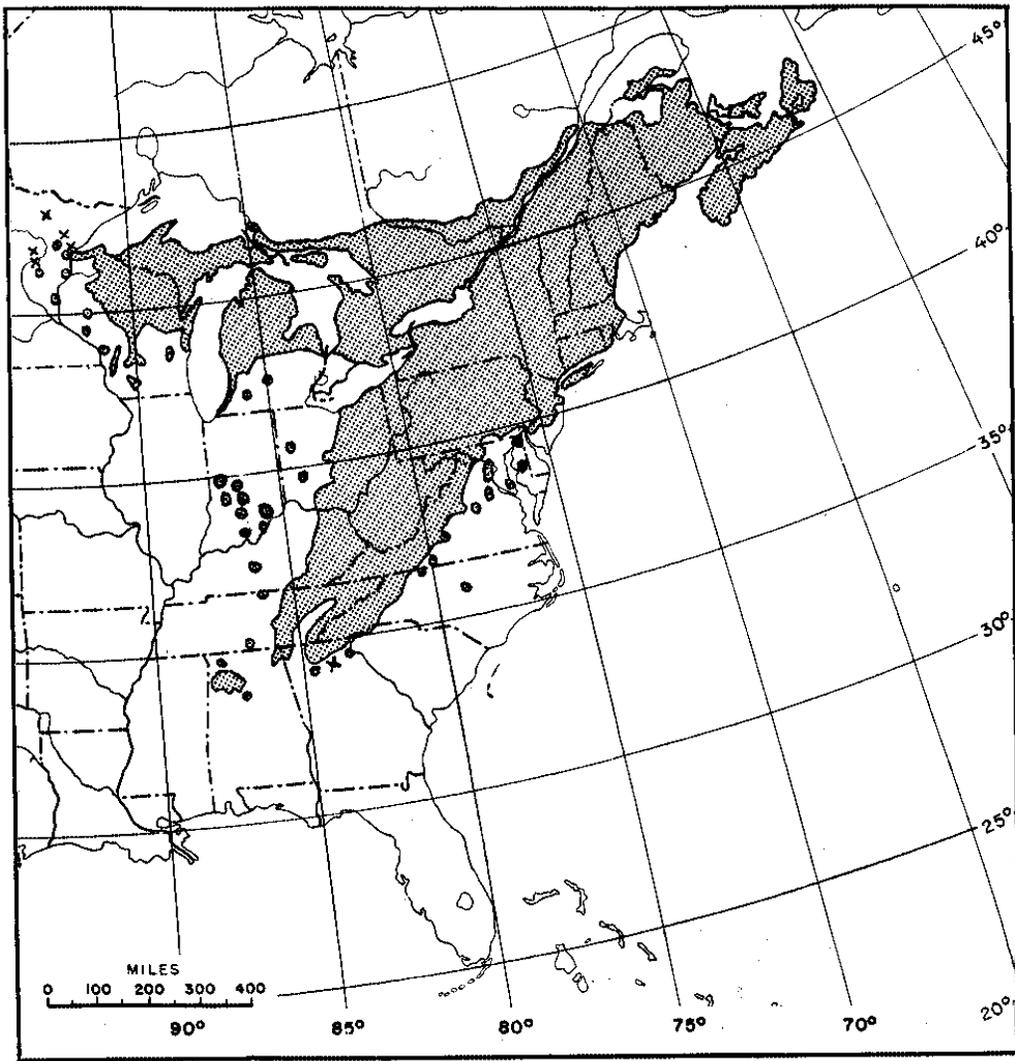


Figure 5: Native range of eastern hemlock (Brisbin 1970).

Eastern hemlocks (hereafter “hemlock”) are noted for their small needles as well as having the smallest cones in the pine family. Needles are typically 0.5-cm wide and .85 to 1.85-cm long, and are flat and narrow. Cones are only 1.25 to 2-cm long (Brisbin 1970). Needles and cones normally grow in very dense clusters. Understory in a dense stand of hemlock is less likely to develop due to the lack of penetrating solar radiation through the canopy, however, low lying shrubs, herbs, and mosses are able to grow (Burns and Honkala 1990).

The hemlock root system is typically defined by site conditions. When high soil horizons are moist, the root system is shallow, but deep root systems have also been observed. Most typically, hemlocks are found in lower lying, shaded areas, where root systems are shallow, making the trees susceptible to windfall and damage from fire (Burns and Honkala 1990, Carey 1993). There is no one soil type requirement for this species, but soils are generally well drained and moist to very moist (Yamasaki 2000). Hemlocks grow in cool humid climates with winter temperatures averaging below freezing ( $-12^{\circ}\text{C}$ ) and summer temperatures moderate ( $16^{\circ}\text{C}$ ) in northern states, while in the southern extent, temperatures are generally much warmer (winter averages of  $6^{\circ}\text{C}$ ) (Burns and Honkala 1990, Carey 1993). Precipitation is highly spatially variable within the hemlock’s range. Values range from 75 cm in the New England States to 150 cm in Southern Appalachia and 1000 + cm for Lake States (Yamasaki 2000, Burns and Honkala 1990).

Although it has been long understood that hemlock provides a vital habitat for white-tail deer, especially as a source of food in the winter, many other species of birds and mammals also rely on hemlock for food, shelter, and breeding (Anderson and Loucks 1979, Yamasaki 2000). Eastern hemlock habitat can be characterized by a number of unique factors preferred by certain types of birds and mammals such as a generally cooler micro-climate, thermal coverage in winter

months, dense overstory and canopy that reduces forest floor light penetration, increased woody debris thickness and coverage on the forest floor, and cavities in trees that provide roosting, denning, and breeding opportunities for birds and mammals. It is estimated that 96 species of bird and 47 species of small and carnivorous mammals are associated with hemlock habitat, while 8 bird species and 10 mammal species are strongly associated, relying significantly on hemlocks for food, shelter, and reproduction (Yamasaki 2000).



Figure 6: Eastern hemlock branch with young green cones. Source: J.S. Peterson @ USDA-NRCS PLANTS Database (USDA 2011).

### Hemlock Woolly Adelgid

The hemlock woolly adelgid (*Adelges tsugae*) (HWA) is a small aphid-like insect that is native to Japan. The HWA has a polymorphic life-cycle in which it undergoes six different developmental phases; egg, four nymphal instars, and adult, which can include both winged and wingless forms (McClure 1987, McClure 1989). The HWA undergoes most of its developmental phases between June and October, and undergoes two annual generations (Orwig and Foster 1998).

The HWA is believed to have been introduced to the western coast of the United States as early as 1922, but western species of hemlock have natural defenses that allowed them to remain unaffected by the pest (McClure 1987). The first report of HWA on the east coast, where hemlock species are susceptible to the invasive pest, was in the 1950s in the Mid-Atlantic States. The HWA spread north and south quickly since that time (Stadler et al. 2005, Orwig and Foster 1998, McClure 1990). As of 2009, it is estimated that HWA has spread as far south as Georgia, as far north as Maine, and as far west as Tennessee and Kentucky, affecting as much as 50% of all eastern hemlock (Figure 7) (USFS 2011).

Immature nymph and adult adelgids feed on the fluid from the base of hemlock needles, where the needle attaches to branches. A female HWA may lay up to 300 eggs in one generation, and multiple generations may overlap. Through sheer number, the HWA is able to overfeed on hemlocks causing loss of vigor, early drop of needles, and lack of regenerative growth (McClure 1987). The adelgid gets its name from the white fluffy sacs that look like wool that form on nymph crawlers, and remain on the insect's body its entire life (Figure 8) (USFS 2011).

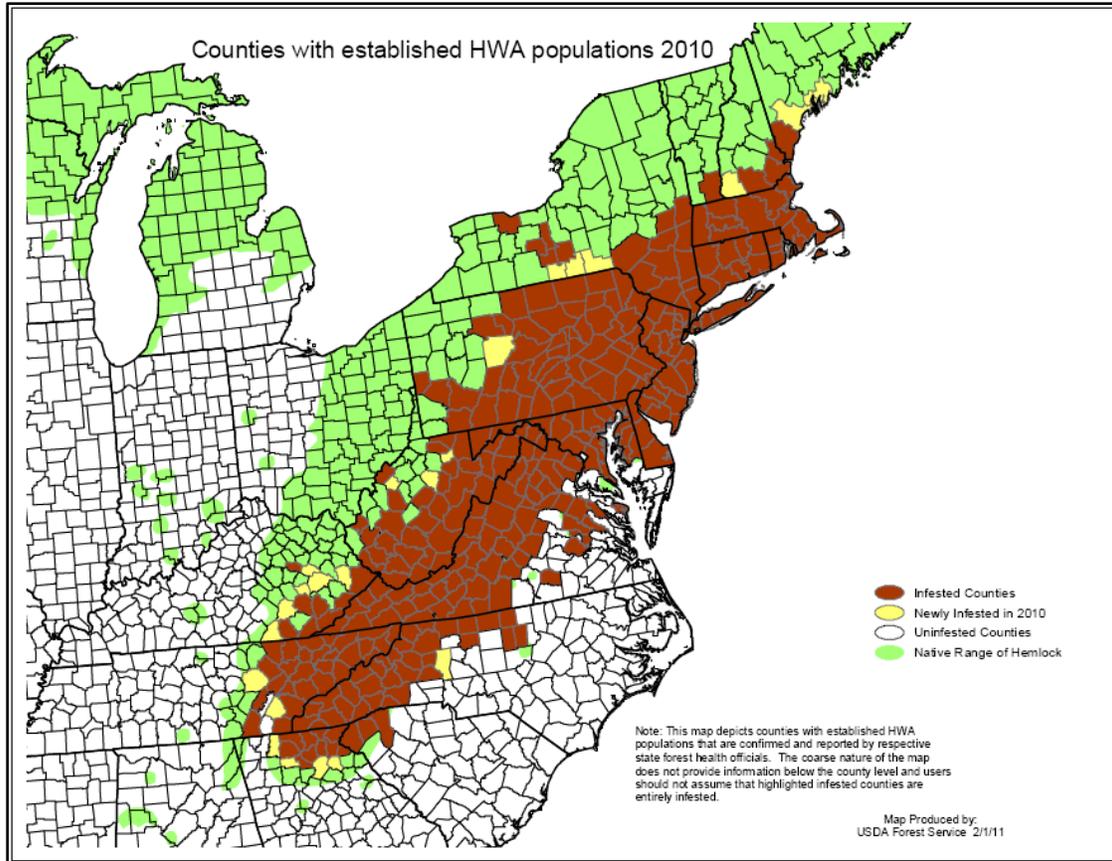


Figure 7: Counties with established and new adelgid populations in 2010 (USFS 2011).

HWA is spread by a number of different processes including wind, birds, and human activity (McClure 1990). McClure (1990) found an even vertical dispersion of the woolly adelgid throughout the canopy, indicating that an infested tree is likely to be covered from base to top. It is predicted that most trees die within 5 – 15 years of infestation, with slightly faster rates found in the south (Stadler et al. 2005, McClure 1990). Research is ongoing as to new forms of treatment, including insecticides, biological controls (natural predators), and hybridization of species (USFS 2011, Orwig and Foster 1998, Orwig et al. 2002).



Figure 8: Woolly sacs found on a hemlock branch (USFS 2011).

### Effects of Hemlock Decline

Unlike other infestations, diseases, or blights seen previously, HWA has the ability to wipe out entire stands of hemlock within a few years, regardless of age (Orwig and Foster 1998). As hemlocks continue to die along the east coast, a number of environmental changes occur. Hemlocks exert a strong control over their stand characteristics, including microclimate, soil characteristics, and habitat (Orwig and Foster 1998, Yamasaki 2000). As hemlocks die, animal species that once relied on hemlocks for food and shelter will have to adapt to new tree species, emigrate, or decline in population. A reduction in the diversity and change in spatial pattern of many plant and animal species is expected as hemlock habitat declines (Yamasaki 2000). Hemlocks create shade with relatively low and dense bottom branches that maintain a cool and moist microclimate. Hemlock seeds are very sensitive to changes in temperature and moisture,

and with increased light penetration to the forest floor due to HWA damage, the moisture of soils and air decreases and temperature increases. Increased solar radiation causes the number of new hemlock saplings and rate of germination to decline (Burns and Honkala 1990, Stadler et al. 2005, Orwig and Foster 1998). Rain through fall also increases in areas of HWA damage, resulting in erosion, especially along banks of creeks and rivers on which hemlocks grow. Changes in stream temperature, pH, discharge, and available photosynthetic light have significant ecological effects on riparian ecosystems (Roberts, Tankersley Jr and Orvis 2009). Species succession occurs rapidly both for new dominant canopy species and understory vegetation. Forest characteristics also change in favor of species that prefer more sun exposure and are more tolerate to warmer and drier conditions (Stadler et al. 2005, Orwig et al. 2002).

#### Ground Measurements of Leaf Area Index

Leaf Area Index (LAI) is the ratio of upper leaf surface of vegetation divided by the surface area of the ground from which the vegetation grow (Watson 1947). LAI is a dimensionless value that ranges from 0 (no vegetation cover) to 6 (dense vegetation cover). Field data collection methods of LAI can be divided into two categories: direct and indirect methods. Direct methods involve either the collection of all fallen leaf litter over a given time period, or the physical removal of all vegetation from a given area. With both methods, the actual upper surface area of leaves is measured, and an LAI value is directly derived. Indirect methods, such as those employed in this study, include categories of measuring leaf area without the collection or removal of leaves from a site. These methods can include contact, such as the use of plum lines and point quadrants, and non-contact methods such as hemispherical photography (Jonckheere et al. 2004).

Hemispherical photography as a ground measure of LAI has both advantages and disadvantages. The capture of photographs in the field is quicker and simpler than other methods, and provides permanent visual reference to the light conditions, species composition, canopy density, age and variation, and distribution of canopy gaps for a given site (Jonckheere et al. 2004). Analysis of photographs, however, is often time consuming, and myriad methods and equations with which to process photographs often leads to disparate and unrepeatable results. One of the first steps to begin analysis of hemispherical photographs is the thresholding of images: deciding what constitutes vegetation and sky in the image based on grey values in black and white or red, green, and blue values in a color image. Image thresholding alone constitutes its own area of research – with one study citing over 40 different methods of thresholding images (Jonckheere et al. 2005).

Calculation of LAI from optical methods is an inexact method due to the inclusion of woody debris, stems, and other plant parts in the calculation of a photograph. Previous studies have labeled LAI from non-destructive methods as  $LAI_{\text{effective}}$  or Plant Area Index, since it is not possible to truly know the measured LAI. It would also be incorrect to mask or remove woody debris from an image during processing, since no assumptions can be made about the presence of leaves behind woody debris in an image (Weiss 2010). Due to this, LAI from hemispherical photographs has been found to generally underestimate the true values of LAI, when destructive and non-destructive methods are tested side-by-side (Chen 1996). Therefore, in this study, values of LAI presented are calculated as *effective* measures of LAI.

In some instances, a clumping factor can be calculated and applied as a conversion factor to more closely calculate true LAI, but the derivation of clumping factors is especially difficult in needle-leaf forests (Chen 1996, Riaño et al. 2004, Weiss et al. 2004). Both the clumping of leaf

elements and the distribution of branches in a site can lead to miscalculations of LAI (Chen and Black 1991). Two equations used in this study have a clumping factor applied to estimate true LAI values.

### LiDAR

New remote sensing techniques augment and expand our ability to study the world around us. Light Detection and Ranging (LiDAR), for example, is a relatively new and unique form of remote sensing that provides data that were not previously available with traditional photographic techniques (Wehr and Lohr 1999).

LiDAR emits a laser light pulse from a transmitting sensor, and receives return pulses that are reflected off a target back to the receiving sensor. LiDAR sensors can be terrestrial (ground based), aircraft based (fixed or rotor wing), or satellite based. All airborne laser scanning (ALS) systems have three main components: the laser ranging unit, the opto-mechanical scanner, and the control and processing unit (Wehr and Lohr 1999). Based on the known travelling time of a light pulse, the LiDAR processing unit is able to calculate the total distance from the sensor to the object, equaling  $1/2$  the total time from transmission to receiving (Wehr and Lohr 1999, Lim et al. 2003). The laser ranging unit emits a continuous wave or a laser pulse, which is directed through a scanning mechanism towards the target. The result of the scanner and the forward motion of the aircraft results in near parallel zigzag scanning lines; however, the resulting scan lines are dependent on the type of scanning mechanism and the speed of the aircraft (Figure 9). After the laser pulse is reflected off an object, the reflected laser is received by a separate sensor and processed. A position and orientation system (POS), which is composed of an inertial measurement unit (IMU) and a differential GPS (DGPS), correctly orients the returning laser pulses in space. The IMU records the roll, pitch and yaw of the aircraft while the

DGPS unit records the exact location of the LiDAR system. This information, combined with the ranges and scan angles, sensor calibration and mounting parameters, enables accurate positioning of each laser pulse in space.

One of the advantages of LiDAR is the ability to achieve high horizontal and vertical accuracies. Generally, current vertical accuracies range from approximately 0.1 to 0.3 m, while horizontal accuracies can be sub-meter to 3 m, depending on the sensor. Advances in LiDAR acquisition and post-processing have begun to increase accuracies, with some new scanners and techniques providing sub-centimeter accuracy (Ananda Fowler, Riegl USA, personal correspondence).

There are two main types of ALS systems used in forest applications (Lim et al. 2003). The first type of laser is called continuous wave (CW) laser or full waveform lasers. Examples of this type of system are SLICER (Scanning LiDAR Imager of Canopies by Echo Recovery) and VCL (Vegetation Canopy LiDAR Mission) (Wehr and Lohr 1999, Dubayah 2000, Lefsky et al. 1999). This type of ALS emits a continuous laser wave that records the intensity of returns (Figure 10). The return intensity is recorded continuously, so that the intensity is near zero when travelling through space, but high when reflecting off a solid object.

The second type of system is a discrete return system, where individual laser pulses are emitted (currently up to 150,000 per second). Early versions of these systems operated at approximately 400 pulses per second and only were able to record one or two returns per laser pulse (i.e., first and/or last returns), however, most modern systems are able to record up to five returns for each laser pulse. Currently, this is the most common commercial LiDAR system. The comparisons between full waveform and discrete return LiDAR systems can be seen in Figure 10.

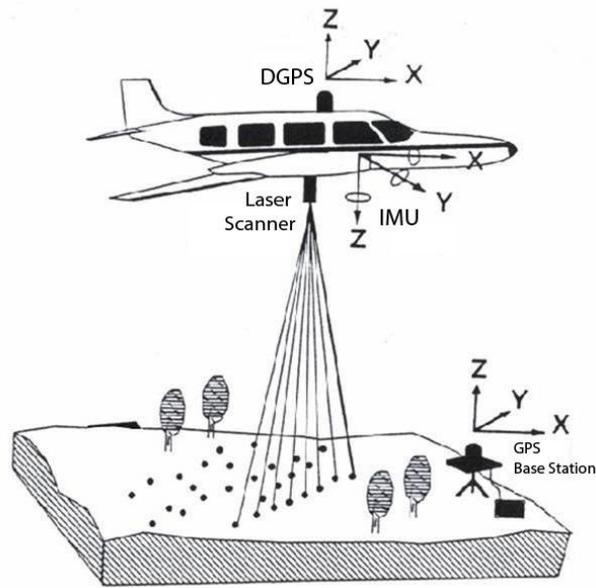


Figure 9: An ALS system with laser scanner, Differential Global Position System, and Inertial Measurement Unit (Lim et al. 2003).

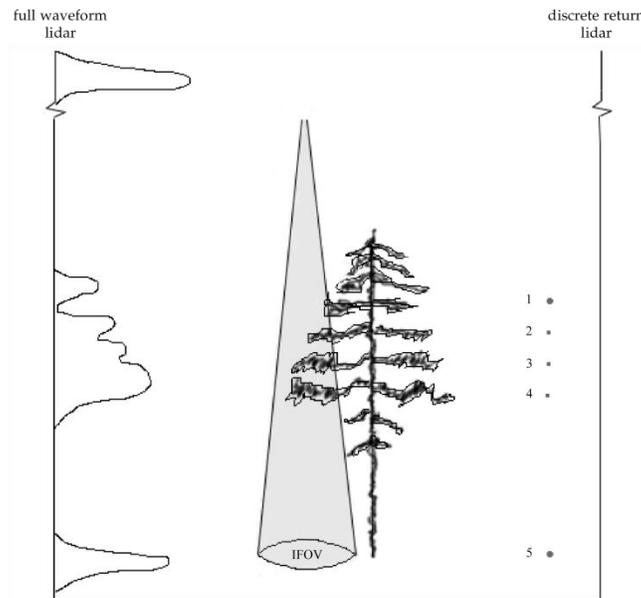


Figure 10: The difference between full waveform and discrete return LiDAR. The cone in the middle represents a single LiDAR pulse. The numbers on the right indicate the discrete return number. Modified (Lim et al. 2003).

A number of factors, including flying height, the field of view (FOV), and instantaneous field of view (IFOV) (i.e., the laser footprint on the ground) determine the density of sampling of LiDAR points, as well as the total area to be covered in a single flight line (Dubayah 2000, Wehr and Lohr 1999). Generally, LiDAR systems are divided into two different classes: small footprint (5 – 30 cm) and large footprint (up to 25 m). For an in-depth discussion of the advantages and disadvantages of large and small footprint LiDAR systems, see Dubayah (2000). Small footprint LiDAR is better adapted for finer scale analysis, as in this study. Large footprint systems cannot always capture the same high resolution detail needed to analyze 20-m diameter or smaller area sites.

### LiDAR Applications in Forest Studies

LiDAR systems are well suited to ecological applications including assessment of forest structure (Lim et al. 2003, Dubayah 2000, Lefsky et al. 2002, Nelson, Krabill and Maclean 1984). LiDAR can directly measure and estimate forest characteristics such as canopy height, canopy density, subcanopy topography, ground elevation, and vertical distribution of objects between canopy and ground (Dubayah 2000). LiDAR measurement are also correlated with other forest characteristics such as crown shape, roughness, tree density/spacing, basal area, biomass, and LAI (Coops et al. 2007, Nelson et al. 1984, Nelson, Krabill and Tonelli 1988, Popescu, Wynne and Nelson 2003, Lefsky et al. 1999, Kane et al. 2010, Means et al. 1999).

Some of the earliest uses for ALS systems were coastal topographic and bathymetric applications, but in the early 1980s, those scanners were adapted for ecological uses, especially for characterization of forests. Nelson (1984), for example, used a full waveform bathymetric laser system in profiling mode to acquire tree heights and canopy density, and compare those values to photogrammetric methods. The laser data were collected over a flight line 14-km long,

but only 0.75-m wide (the IFOV of the scanner) at 400 points/sec (approximately 0.25-m post spacing). Only 2 returns per pulse were collected. They found laser height measurements could estimate tree heights within approximately 1.0 m of traditional photogrammetric methods, and they were able to sample heights at a greater rate. They observed healthy canopy was more likely to return values approximating the top of the canopy, while defoliated canopy (in their case, due to gypsy moth infestation) returns were more likely to have been hits from below the canopy surface. He also found they could approximate canopy density better than traditional photogrammetric methods with greater certainty. His study is one that laid the groundwork for all future work in ALS for forest characteristics by postulating that LiDAR data could accurately be correlated with other forest metrics, including biomass and timber volume (Nelson et al. 1988).

Prior to the mid-1990s, most ALS studies were accomplished in profiling mode (Næsset 1997). This greatly limited the capacity of researchers to assess large areas of forest without the need for extensive field or photogrammetric based work, which was expensive and time consuming. These early studies also suffered from a lack of high accuracy location information. However, as DGPS, IMU and laser scanner technology advanced, researchers had greater access to commercially available scanning systems that covered larger areas with fewer flight lines. The advancement of computing technologies, especially post-processing software for laser data, aided in the wider use of LiDAR data (Næsset 1997, Wehr and Lohr 1999). Increased data storage capabilities also allowed researchers in the late 1990s to begin collecting and processing multiple return data for both full waveform and discrete return data.

In spite of initial success in applications of LiDAR in forest studies, LiDAR derived mean tree height generally underestimated true mean heights (Magnussen and Boudewyn 1998,

Næsset 1997). This problem has been approached in a number of different ways with varying degrees of success. The basic assertion is that it is unlikely for a laser pulse to: a) hit the highest parts of the canopy, such as tree crown tops; and b) regardless of where laser pulses hit, they penetrate canopy by up to a few meters before being returned. Næsset (1997) found that laser data generally underestimated ground truth heights by 4.1 – 5.5-m. However, by using a grid sampling method of the highest points per cell, and by increasing or decreasing the cell size, they were able to reduce the bias to -0.4 – 1.9 m. Other, more mathematically complex methods have also been proposed with varying degrees of successful application (Magnussen and Boudewyn 1998).

In addition to forest height, LAI is another important metric that LiDAR can measure with increasing levels of accuracy. As observed in previous work, one limitation of photogrammetric methods for estimation of LAI is saturation at high values, and noise caused by variability in reflectance values (Chen and Cihlar 1996, Riaño et al. 2004). Using high sampling (9.3 points/m<sup>2</sup>), small footprint (0.45-m diameter) discrete return data, Riaño et al. (2004) showed they could achieve very high coefficients of determination by varying the size of the sampled LiDAR data for a given hemispherical measure of LAI in two contrasting forests, achieving R<sup>2</sup> values of about 0.9 and 0.5 for oak and pine, respectively. Other studies have also found that it is possible to explain LAI accurately through the use of LiDAR datasets, including Lefsky et al.'s (1999) use of SLICER full waveform laser to estimate LAI and explain 75% of the LAI variability in a Douglas-fir Western hemlock forest using a three-dimensional canopy volume method (CVM).

Being able to recognize subtle differences in canopy structure between forest stands is an essential part in describing the effects of a fine scale and spatially heterogeneous disturbance on

a given area. Canopy structure changes dramatically during and after stand disturbances, and being able to quantify those changes depends on the ability to show a measurable difference in stand structure (Lefsky et al. 1999). Solberg et al. (2004) tracked the progress of a severe pine sawfly attack on scots pine forests using LiDAR derived LAI measurements and found that correlation coefficients were strong: 0.94 with LAI-2000 and 0.87 with hemispherical photography. The use LiDAR before and after a disturbance has not been widely studied since multi-temporal LiDAR is relatively unavailable. Until the costs of LiDAR acquisition is low enough for monitoring on a regular basis, alternative methods for detecting disturbance and defoliation in canopy is required. This study attempts to explain effects of invasive insect defoliation on hemlock forest structure using methods described in Chapter 3 of this thesis.

## CHAPTER 3

### METHODOLOGY

The methodology for this study can be divided into four different categories: field data collection methods, field data processing, LiDAR data processing, and field and LiDAR data analysis. Each methodology required its own preparation and execution, and in some instances, previously proposed methods were not appropriate and new methods were adapted to achieve results.

#### Field Data Collection

Field data were collected at 36 sites during the months of April and May, 2011. Before field data collection was carried out, selection of appropriate field sites to meet the proposed objective criteria for field sites was necessary. This proved to be difficult and posed a large obstacle to the collection of field data. The first attempt at identifying field sites included the use of a GIS to overlay both USFS Hemlock Conservations Areas (HCAs) and “pure” hemlock vegetation cover classification data on both USDA NAIP 2010 imagery and custom acquired high resolution orthoimage tiles collected in tandem with the LiDAR data as part of a USGS-funded grant for researches at Gainesville State College (Sharma and Panda, personal communication). Due to the spectral similarities between species of pine and hemlock in the true color NAIP images, image interpretation provided limited utility in identifying hemlock stands. Three separate trips were made to HCAs throughout the central and northeastern part of CNF to ground truth possible study sites. In all cases, stands were typically pine dominated overstory with mixed hemlock as a subdominant species.

### *Field Site Selection*

With assistance from University of Georgia Warnell School of Forestry and Natural Resources researchers involved in insect release and biological control of HWA throughout CNF, five maps were acquired with treatment sites. In each area, previous research sites involving HWA treatment were only considered as guides from which to begin a search to find other appropriate sites. In some cases, but not all, sites for this study coincided with sites selected for HWA control. Actual field site selection was based on a number of factors. Each site represented at least 50 percent hemlock coverage in the main canopy. This was assessed visually from the ground for each site and based on percentage of main canopy species stem count. In order to focus on HWA disturbance to canopy trees, understory was considered to be undesirable, and in the majority of sites, there was little to no presence of an understory or low lying vegetation. Proximity to interfering features, such as power lines, houses, cliffs, and rivers was considered undesirable in site selection; however, as discussed in the literature review of eastern hemlock, these trees show a preference for steep, rocky slopes near water. The majority of chosen sites were flat, but in some instances, sloped sites were selected due to the preference for high hemlock canopy percentage. In all cases where sites were selected near water or roads, care was taken to be outside of a buffer from roads or water of approximately an additional 5 m to ensure that both LiDAR and georeferenced ground data would not overlap with water or road features. Due to the limited distribution of known hemlock habitat in the CNF, sites were not selected in a systematic or grid pattern. Alternatively, sites were selected where criteria were met and spatial autocorrelation between sites was not considered. Of these five research areas, four were chosen for field survey in this study. The geographic coordinates for these areas were loaded onto a Garmin Vista HCx handheld GPS unit with Wide Area Augmentation System

(WAAS) correction to approximately  $\pm 3$  m. This unit was then used as a reference guide when looking for sites.

Once a site was selected, a research protocol was followed for the collection of all data at each site to ensure consistency. While details are provided below, this included first collecting a high accuracy GPS point location, then collecting a spherical site survey image, and finally collecting the hemispherical canopy image. A field survey, which included a total count of all main canopy tree stems within 10 m of the survey point, was also collected.

#### *Collection of GPS Location*

GPS points for each site were collected for each site using a Trimble Pathfinder ProHX GNSS receiver (Trimble 2011b) and a Trimble Nomad handheld field computer ((Trimble 2011a). GPS points were collected using real-time WAAS corrections. The averages of the reported accuracies were  $\pm 1.5$ -m Horizontal Positional Error (HPE) and  $\pm 2.4$ -m Vertical Positional Error (VPE) with an average Position Dilution of Precision (PDOP) of approximately 5 (Table 2). The locations were not post-processed from the Trimble unit since the real-time accuracy meets or exceeds that of many other LiDAR canopy studies.

The GNSS receiver and field computer were mounted on a 2-m tall survey pole and connected via Bluetooth. Each point was averaged 30 times per site before recording a location with a target maximum PDOP of 6 and a maximum allowable PDOP of 7. Locations data was recorded both from Trimble GPS Controller and ArcPad 7.0 software on the handheld field computer.

Table 2: HPE, VPE, and PDOP for all study sites (units in  $\pm$  meters).

HCA ID	HPE	VPE	PDOP	HCA ID	HPE	VPE	PDOP
<b>145-01</b>	0.78	2.5	4.94	<b>29-11</b>	2.6	3.1	4.8
<b>145-02</b>	1.1	2.5	5.34	<b>29-12</b>	3	3.2	5.8
<b>145-03</b>	1.2	2.1	3.51	<b>71-01</b>	2.1	3.2	6.5
<b>145-04</b>	1.5	2.6	5.71	<b>71-02</b>	3	3.5	5.81
<b>145-05</b>	1.9	2.7	6.8	<b>71-03</b>	1.3	3.1	4.66
<b>145-06</b>	1.8	3.7	6.88	<b>71-04</b>	1.8	2.4	5.69
<b>145-07</b>	0.8	1.4	3.07	<b>71-05</b>	0.76	1.2	2.98
<b>145-08</b>	2	3.1	6.12	<b>71-06</b>	0.76	1.8	3.79
<b>29-01</b>	0.96	1.9	4.1	<b>71-07</b>	0.91	2.2	3.77
<b>29-02</b>	2.1	2	5.5	<b>71-08</b>	0.69	1.5	4.89
<b>29-03</b>	1.9	2.4	5.88	<b>71-09</b>	1.5	3.6	7.2
<b>29-04</b>	1.1	2.3	4.8	<b>71-10</b>	0.63	0.99	2.26
<b>29-05</b>	1.7	1.9	4.77	<b>72-01</b>	1.6	2.2	5.8
<b>29-06</b>	0.89	2	2.88	<b>72-02</b>	1.7	2.4	3.69
<b>29-07</b>	1.2	1.9	4.28	<b>72-03</b>	1	2.9	5.87
<b>29-08</b>	1.9	2.3	5.6	<b>72-04</b>	0.94	3.2	6.5
<b>29-09</b>	2	2.3	5.6	<b>72-05</b>	2.2	2.7	6.58
<b>29-10</b>	1.2	1.5	4.6	<b>72-06</b>	0.87	1.1	5.47

### *Collection of Field Imagery*

A Nikon CoolPix 8700 camera body was used to collect all spherical and hemispherical photography. The system, assembled by IPIX (IPIX 2011) is designed to acquire two images that are later knitted together to create 360° immersive photography – imaging the entire sphere of area around the camera. The CoolPix 8700 has 8.0 million effective pixels. All images were collected as JPEG files on the fine quality setting with an image resolution of 3264 x 2488 pixels (highest available). The camera body lens was set at 35mm. All other settings were pre-programmed from IPIX; therefore, shutter speed, ISO, and aperture settings were set to automatic, and flash was turned off. A Nikon FC-E9 0.2x fisheye lens convertor was mounted to the camera body with a UR-E12 lens adaptor. This fisheye lens provides approximately 183° to

190° view angle for similar camera models, though the exact view angle is not specified for the CoolPix 8700. The fisheye lens reduces the focal length of the camera's lens to x0.2.

For all study sites, a pair of horizontal images, level with the ground, were taken that were later knitted together to create the spherical immersive image. This process uses a special tripod and camera mount that allows the camera and lens to rotate. The mount has notches at 0° and 180°. Two images were taken facing exactly opposite of each other. Photo-documenting each site using immersive photography allows for future study of changing conditions of hemlock canopy and the surround forest, as the forest floor is visible in all spherical images. This mosaic also allows a second look at certain site conditions that may not have been originally documented while in the field. Two sets of IPIX immersive spherical images were collected at each site for site documentation.

After the spherical photography was collected, the camera was remounted vertically to a tripod that allowed the camera to face directly up at the canopy for collection of hemispherical canopy images which were used to calculate forest metrics. The fisheye camera system was first oriented so that the top of the camera faced north and then leveled using a torpedo bubble level. Three images were collected per site. All photographs were taken in clear, blue sky conditions. Blue sky conditions are preferred for this study over other conditions (such as overcast) because vegetation to sky contrast is higher during clear, well lit conditions, than in cloudy or poorly lit conditions. However, clear sky conditions introduce the sun as an obscuring factor when thresholding and interpreting images. Caution was used when collecting these images to ensure that exposed sun did not interfere with image collection, but not all images were free from sun. This can both introduce highly overexposed areas of the image creating a halo effect, as well as

overexposed vegetation and tree trunks. This issue is partially handled by capturing three images per site, and partially handled during image processing, which will be explained later.

### *Site Description and Other Variables*

A basic site description was recorded that included the perceived HWA site status from the ground, as well as a main canopy stem count and hemlock stem count. The Health Status Ranking (HSr) was a subjective value assigned to a site that defined whether it was: (1) healthy; (2) moderately healthy; or (3) unhealthy.

An assessment was made for each site about the damage to the surrounding hemlock trees due to HWA. A site with limited visible damage was given a HSr of 1, while a site with extensive damage, dead branches, and damage that extended upward on the tree trunk was given a HSr 3. A stem count was conducted within a 10-m radius from the center of the site. This included counting all stems that were determined to be main canopy, generally taller than 4 – 5 m without weight given to diameter at breast height. Hemlocks only were then counted, and a percentage was calculated of hemlock to total stems to give a rough estimate of the percentage of hemlocks in the main canopy for a given site. The values collected during the field site can be seen in Table 3.

Table 3: Health Status Rank (HSr), Stem Count, Hemlock Stem Count, and Hemlock Percentage based on stem count for all study sites.

HCA ID	HSr	Stem Count	Hemlock Count	Hemlock Percent	HCA ID	HSr	Stem Count	Hemlock Count	Hemlock Percent
145-01	2	59	30	51	29-11	1	72	56	78
145-02	3	46	25	54	29-12	1	28	18	64
145-03	2	37	30	81	71-01	2	68	52	76
145-04	3	43	24	56	71-02	3	60	43	72
145-05	2	50	35	70	71-03	1	45	27	60
145-06	2	43	31	72	71-04	3	24	12	50
145-07	1	61	42	69	71-05	2	59	35	59
145-08	1	40	21	53	71-06	3	52	34	65
29-01	2	50	31	62	71-07	1	38	21	55
29-02	2	30	20	67	71-08	2	41	25	61
29-03	2	39	28	72	71-09	2	37	22	60
29-04	1	61	51	84	71-10	3	33	17	52
29-05	2	100	86	86	72-01	1	48	42	88
29-06	3	45	38	84	72-02	1	53	48	91
29-07	3	34	27	79	72-03	3	58	44	76
29-08	2	50	35	70	72-04	3	48	24	50
29-09	2	49	36	74	72-05	3	80	60	75
29-10	3	41	26	63	72-06	1	37	24	65

### Field Data Processing

Once all images were collected in the field, they had to be corrected for lens distortion and undergo threshold processing. Calculations could then be made to determine measures of forest structure. The Can-Eye V6.1 (INRA 2011) free software package was used to process all hemispherical field images. This software was developed by the French National Institute of Agronomical Research and designed to calculate the fraction of soil covered by vegetation, fCover, Leaf Area Index (LAI), (both effective and true), Average Leaf Inclination, and a host of other canopy metrics.

### *Lens Correction and Parameter Selection*

Prior to any image analysis, a lens correction procedure was applied to ensure correct calculation of all metrics from the hemispherical photography. The procedure required two separate steps to calculate a third order polynomial lens correction equation for the CoolPix + fisheye lens system. The first step was the calculation of the optical center of the camera sensor. In a perfect system, the optical center of an image is exactly  $\frac{1}{2}$  the vertical pixels and  $\frac{1}{2}$  the horizontal pixels, but due to flaws in the camera sensor and lens, the optical center is not typically the center X and Y pixels of a digital image. To determine the optical center of the picture, three holes were put into the lens cap. A series of 17 images were taken with the lens cap on, and the lens cap was rotated counter clockwise between each image. The X and Y values for each lens hole in each image were input to an Excel spreadsheet provided with the Can-Eye software.

The second step in lens correction was the characterization of the projection function. This was to determine the radial distortion of the lens as one approaches the edge of the image from the optical center. To complete this process, three 1-m sticks were used and attached together in an open sided rectangle measuring 50-cm on the parallel sides and 30-cm on the open side (Figure 11).

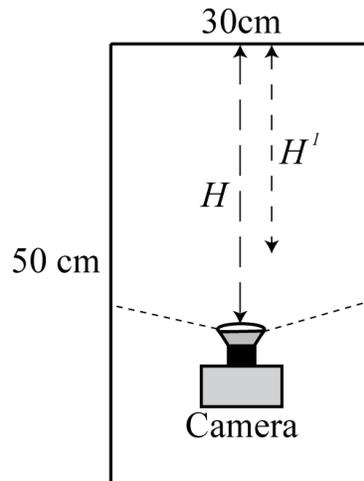


Figure 11: Camera calibration was an important step to fixing lens distortion. Two images were taken at distances  $H$  and  $H'$  from the meter sticks (the open ended rectangle) and X and Y pixel values of centimeter readings on the meter sticks were extracted with imaging software and input to a spreadsheet.

Two images were taken at distances of  $H$  and  $H'$  from just inside the rectangle towards the 30-cm open side so that the centimeter readings of the meter sticks were visible in the images. The images were loaded into ERDAS Imagine, and the X and Y image pixel values of centimeter readings from the meter sticks in the images were recorded in the provided Can-Eye spreadsheet (Figure 12). Can-Eye uses the X Y image coordinates of the visible centimeter readings from meter sticks to calculate the distortion in the lens. The optical center and the projection photos are required to calculate a third order polynomial correction which was applied to all photos prior to data extraction.

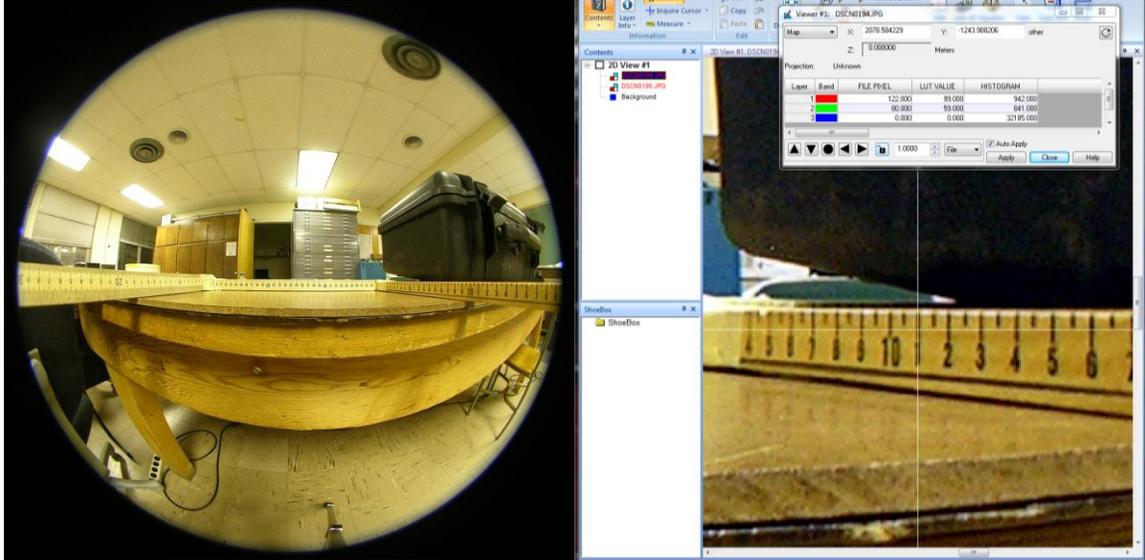


Figure 12: The left image is the lens correction photograph at distance  $H$  and the right image shows the X Y pixel value extraction from the centimeter readings on the right meter stick.

All hemispherical images were processed using the same parameter functions. These parameters were saved in a master file that was applied to each photo prior to processing (Table 4). The zenith and azimuth angular resolution defines the resolution at which LAI is modeled. The Circle of Interest angle defines the area of the image that is used in calculating all variables. This value could range as high as  $90^\circ$ , but in order to ensure that areas of the ground in sloped study sites were not included in the images,  $60^\circ$  was chosen as a value that included only vegetation and sky.

Table 4: Can-Eye Parameter settings were saved in a master file and used to calibrate each hemispherical canopy image prior to calculating LAI or fCover.

Parameter	Value Setting
Zenith Angular Resolution	2.5°
Circle of Interest	60°
Integration Domain for fCover	0 - 10°
Azimuth Angular Resolution	2.5°
Subsampling Factor	1

The Circle of Interest angle is measured from nadir, or directly upward from the lens. Imagine a cone extending from the lens of the camera upward towards the canopy. Now imagine a cylinder that intersects the outer edge of that cone at the top of the canopy and extends back towards the ground. The area of the circle where the cylinder intersects with the ground is the ground footprint area. A larger Circle of Interest angle would create a larger cone, sample more canopy, and produced a larger ground footprint. Therefore, a 90° angle would be horizontal to the ground and may actually include ground *in* the image. 60° was selected as being large enough to ensure coverage of the 20-m site as defined by the ground footprint, and small enough to include only those areas of detail desired.

#### *Hemispherical Image Processing*

Three hemispherical photographs were used to derive all results for each site. While Can-Eye recommends using up to eight images, three were chosen in favor of field and processing time. Can-Eye provides a pre-classification gamma stretch and image masking which aids in both the automatic and manual classification of images. Masking allowed the removal areas of the photograph that were not appropriate for classification or calculation (i.e., an obstruction in the image such as a person). No photographs in this study required masking. Can-Eye used a

modified supervised classification process to threshold vegetation and sky classes for each image. After applying a gamma stretch to the images, there were three options for automatic classification: Greenness Index, Brownness Index, and Brightness Index. These values could be adjusted between 0 and 1 to threshold images. Once appropriate values were selected, Can-Eye provided manual tools for image classification: a pixel based supervised reclassification, and a manual polygon delineation tool for reclassifying areas that would not properly classify using supervised methods. This pixel based classification tool was useful to select values that were mixed pixels, where a prior knowledge helped to correctly classify the pixel values (i.e., dark “blue” that was vegetation, but classified as sky). The polygon tool was used in areas where overexposure due to sunlight caused vegetation to not be classified as such, and could be used on a case-by-case basis. In all cases, polygon delineation and supervised pixel reclassification resulted in satisfactory results where sun exposure correction was needed. Each set of images for each site underwent the same process in Can-Eye prior to variable calculation. An example of the final result of a Can-Eye classification can be seen in Figure 13.

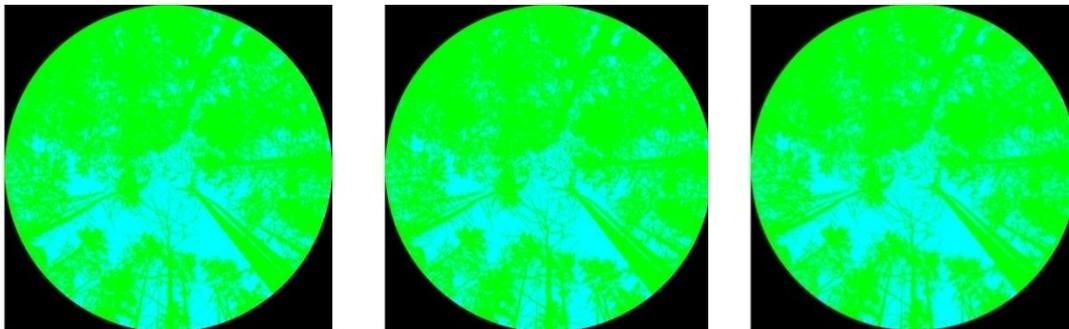


Figure 13: Can-Eye classified each image into two classes: vegetation and sky. In this image, green represents the vegetation class and blue represents sky. Average vegetation coverage of all three images was used to calculate fCover and LAI. HCA ID 29-10.

## *LAI and fCover Calculation from Hemispherical Photography*

The Can-Eye software package calculates 7 different LAI estimations. It uses its own calculation, Can-Eye LAI effective, a traditional Miller LAI effective, as well as estimations of the LICOR LAI2000 instrumentation values at three, four and five rings. Beyond calculating effective measures of LAI, as described in Chapter 2, it also calculates two estimations of true LAI: Can-Eye LAI true and Miller LAI true, where a clumping factor is applied in an attempt to correct for underestimation from effective measures. Can-Eye must model the gap fraction derived as an average of all three images used per site. The following equations expand on those outlined in the LAI section of the literature review.

For this study, only the Can-Eye V6.1 equations were used, since the authors of the software believe the V6.1 equations better calculate LAI than the V5.1 equations used in previous versions of the software (Weiss 2010). The Can-Eye LAI value is based on a Poisson model ( $P_0$ ) which estimates LAI from gap fraction instead of contact frequency. Therefore, Can-Eye Effective LAI can be expressed in the following equations (Weiss 2010):

$$J_k = \sqrt{\frac{\sum_{i=1}^{Nb\_Zenith\_Dir} w_i \left( P_o^{LUT(k)}(\theta_i) - P_o^{MES}(\theta_i) \right)^2}{\underbrace{\sigma_{MOD}(P_o^{MES}(\theta_i))}_{\text{First Term}}}} + \left( \frac{LAI^{LUT(k)} - LAI_{P57}^{Eff}}{\underbrace{\sigma_{MOD}(P_o^{MES}(57^\circ))}_{\text{Second Term}}}} \right)^2$$

$$w_i = \frac{NPix_i - Nmask_i}{NPix_i}, \quad \sum_{i=1}^{Nb\_Zenith\_Dir} w_i = 1$$

Equation 1: Can-Eye V6.1 LAI effective calculation.

This estimation is performed by model inversion because it can take into account a wider range of zenith angles as compared to the alternative Miller formula, which is an advantage when the chance of mixed pixels is higher at larger zenith angles. This formula computes a random

combination of LAI (between 0 and 10 in steps of 0.1) and Average Leaf Angle (ALA) (between 10° and 80° in steps of 2°) values to build a database of Look-Up-Table (LUT) values which correspond to gap fraction values. The process then consists of selecting the LUT element value that is closest to the measured  $P_\theta$ . The first term is a weighted RMSE between the measured gap fraction and the LUT value from the database. The weight  $w_i$  accounts for the inclusion of mixed pixels in some zenith angles and therefore would not be representative of the image.

The Miller formula derives LAI from the gap fraction measured in all directions by determining a contact frequency function (Miller 1967). This formula was presented as an alternative to the point quadrat method, where problems may arise in the interpretation of the data obtained. The formula estimates an average foliage density of a plant in terms of the calculated contact frequency.

The LAI2000 LAI calculations are designed to correlate well with LAI readings from the specialized LICOR LAI2000 instrument, which uses the average angular response as provided by the instrument maker and adapts the Miller formula to match the expected readings of 3, 4, and 5 ring readings from the LAI2000 (Weiss 2010).

“True LAI” is related to Effective LAI as expressed by:

$$LAI^{eff} = \hat{C}_o LAI$$

Where  $\hat{C}_o$  is an aggregation or dispersion parameter or clumping index derived from a clumping factor algorithm (Weiss 2010, Lang and Yueqin 1986, Chen and Black 1992):

$$\lambda_o(\theta, ALA^{eff}) = \frac{mean[\log(P_o^{Cell}(\theta))]}{\log[mean(P_o^{Cell}(\theta))]}$$

fCover is defined as the fraction of soil covered by vegetation viewed in the nadir direction, and averaged over all three photos:

$$fCover = 1 - P_o(0)$$

The fCover is integrated over a range of zenith values ( $P_o$ ) because it is not possible to get a value in the exact nadir direction using hemispherical canopy images (Weiss 2010, Weiss et al. 2004). As discussed previously, the parameter range for fCover integration was set to 0 - 10° (Jonckheere et al. 2004, Weiss et al. 2004).

### LiDAR Data Processing

For this study, methods were adapted that only took into account first, last, and single returns from the LiDAR point cloud. While the sensor recorded intermediate returns, these values were filtered from the point cloud prior to variable calculation. This has been common practice in forest LiDAR literature, and the methods and calculations used in this study adapt themselves well to using only first, last, and single returns (Riaño et al. 2004, Morsdorf et al. 2004, Coops et al. 2007, Kane et al. 2010). Intermediate terms present a problem when calculating fCover. When fCover is calculated from LiDAR as the fraction of canopy returns to ground returns, the introduction of hundreds of canopy returns in the form of intermediate values highly overestimates the fCover. LAI calculations typically involve transformation of point return numbers over a given area simply as a percentage. In this study, a calculation that estimates LAI using the interrelationship of first, last, and single canopy returns is used – making the need to calculate LAI with intermediate returns unnecessary. This measure of LAI more closely resembled non-destructive measures of field survey where plum lines were dropped from above the canopy to calculate a contact frequency. In this case, lasers are substituted for plum lines.

Classification, filter, and cropping of the data prior to analysis were essential. Data were received in the LAS V1.2 file format in 1.5km x 1.5 km tiles of raw point data. These data included four original classes: Unclassified (1), Ground (2), Noise (7), and Overlap (12). For

calculation of LAI and fCover, all vendor classes were removed. Using the QCoherent LP360 software suite for ArcGIS (QCoherent 2011), a height classifier was used to classify all points higher than 2 m as vegetation and below 2 m as ground returns.

$$R_{Vegetation} = R_{Total} > 2 \text{ meters}$$

$R_{Vegetation}$  and  $R_{Total}$  denote all vegetation laser returns and total laser returns, respectively.

In all cases, the Noise class was excluded from analysis. Two output datasets were created: one which included overlap, and one that did not. This inclusion is based on the vendor classification of points. Essentially, this inclusion or exclusion is intended to simulate the sampling frequency of LiDAR. In almost all sites for this study, inclusion of overlap doubled the total LiDAR points per site. Two meters was selected as the height above ground level considered above the height of the camera lens for collection of hemispherical canopy photography, and this level provided confidence that ground returns could not be misclassified as vegetation or vice versa. From the original study site point locations, three buffers were created: 10 m, 7.5 m, and 5 m (equating to 20, 15, and 10-m diameter circular polygons) in ArcGIS. These are known as “data traps”, or areas within which raw laser data are collected and calculated (Figure 14). These polygons were used to crop each study site. Using LP360’s Point Cloud Statistics Extractor Tool, classified point counts were exported for each study site at each buffer size with and without overlap.

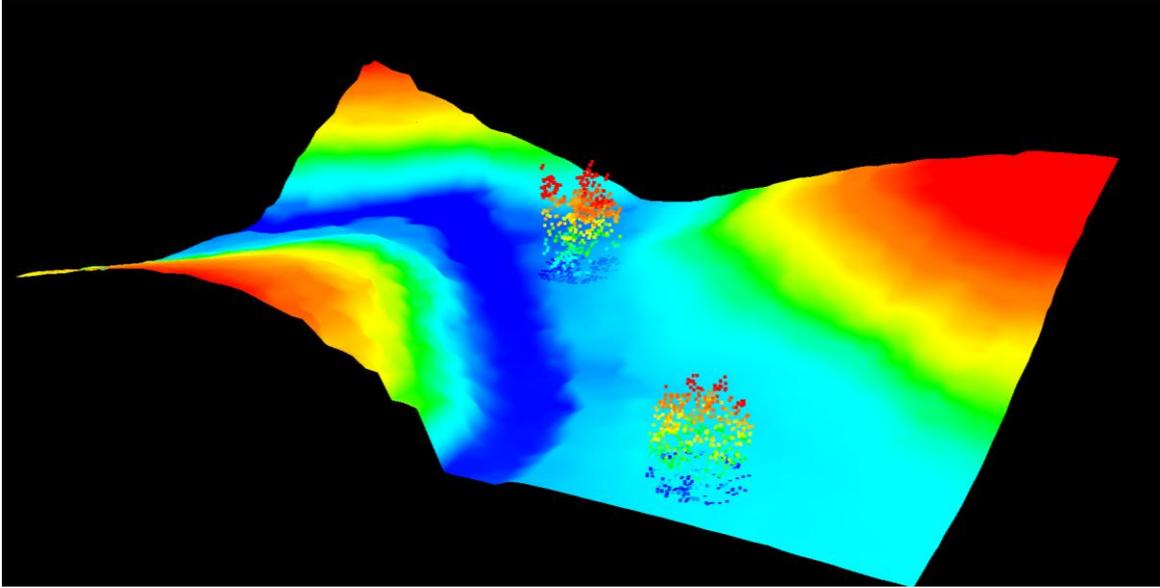


Figure 14: Two 20-m data trap point clouds (29-03 and 29-04) superimposed on a QT Modeler digital elevation model surface.

#### *fCover and LAI Calculation from LiDAR*

Both fCover and LAI variables were calculated from first, last, and single return data at all buffer levels. fCover values are derived by:

$$fCover_{ALS} = \frac{\sum R_{vegetation}}{\sum R_{Total}}$$

Airborne Laser Scanner (ALS) derived fCover, denoted as  $fCover_{ALS}$ , can also be calculated so that  $R_{vegetation}$  and  $R_{Total}$  include only the use of first, last or single returns. These are denoted as  $fCover_{ALS-FR}$ ,  $fCover_{ALS-LR}$ , and  $fCover_{ALS-SR}$ , respectively. It has been previously found that the best estimate  $fCover_{ALS}$  value can be expressed as (Lovell et al. 2003, Morsdorf et al. 2004):

$$fCover_{ALS-FR} > fCover_{ALS} > fCover_{ALS-LR}$$

While LAI cannot be directly calculated from laser derived data, there are proxy variables that can be calculated that have shown strong correlation with hemispherical LAI (Morsdorf et

al. 2004, Riaño et al. 2004). For this study, a LAI proxy value is calculated from a modified version of that found in Morsdorf (2004):

$$LAI_{ALS} = \left( \frac{\sum R_{FR}}{\sum R_{LR} + \sum R_{SR}} \right) * 2$$

In the above equation, all returns are for vegetation only, that is, laser returns greater than 2 m above ground level.  $LAI_{ALS}$  denotes laser derived LAI.

Both  $fCover_{ALS}$  and  $LAI_{ALS}$  were calculated at 20-m, 15-m, and 10-m study site diameters to assess the significance of various correlations with ground based  $fCover$  and LAI due to the collection size of LiDAR.

#### *Above Ground Level Height Estimates*

Using the QT Modeler LiDAR software (Applied-Imagery 2011), Above Ground Level (AGL) values were calculated for all first, last and single return 20-m study site data traps, including and excluding overlap. A 3-m Digital Elevation Model (DEM) was first calculated in the .QTT proprietary file format using only vendor classified ground points for all LiDAR tiles. The import parameters used were QT Modeler recommended settings: Adaptive Triangular with Mean Z Averaging. Hole filling and spike removal algorithms were used to smooth the DEM in QT Modeler. Each 20-m data trap with raw data points was imported to QT Modeler and the height above the DEM surface was calculated in meters for each point. These height values were exported to a .XYZ file with the associated X, Y, Z, and AGL values. All study site ground points were exported to a reference file and used as an error estimator for the AGL calculations (Table 5).

The highest calculated RMSE for a DEM surface was approximately  $\pm .22$  m, with a Z Bias of less than  $\pm .02$  m, giving the calculated AGL values for all raw LiDAR points a high

confidence level. Z Bias is a value that expresses if the DEM surface over or underestimates the raw LiDAR points, and by how much. For example, for HCA 145, QT Modeler overestimates the DEM surface from raw ground point by approximately .02 m. This table indicates that all ground AGL values were within a very tight confidence interval when compared to the ground surface from which all AGL values were calculated.

Table 5: RMSE and Z Bias for AGL Ground calculations for all study sites over their respective tiles.

HCA (All Study Sites)	Tiles	RMSE (m)	Z Bias (m)
145	1257, 1258	0.15	0.02
29	1479, 1522, 1573, 1574	0.22	0.02
71	1234	0.12	0.00

A group of variables called In Canopy Vegetation Fractions (ICVF) were calculated using the AGL values. These values were Low Vegetation, Medium Vegetation, and High Vegetation. A range was calculated for each site from the 2 m AGL level to the mean canopy height (the canopy height estimation). This height was divided into 3 bins, so that each bin represented 1/3 the total height of vegetation points. The points from that bin were extracted and divided by the total vegetation points to derive ICVF. These variables indicate the total fraction of returns seen from the bottom, middle and top 1/3 of the canopy.

#### *Canopy Height Estimation*

To calculate canopy height, the raw LiDAR point cloud was again imported into QT Modeler. This time, the recommended surface settings were used to create a Canopy Surface Model (CSM): Adaptive Triangulation with Z Maximum, 1.0 surface bin smoothing with a 3-m

kernel, and only first returns classified as vegetation were used to create the surface. Hole filling and spike removal algorithms were used, this time more aggressively due to the highly variable nature of surface models over forest canopies. Due to the high sampling rate, a 1-m DEM was created. Both the DEM and CSM were exported into an ESRI ASCII file format. These files were imported and processed through a custom ESRI Model Builder script tool which subtracted the height of the DEM from the CSM and output a Zonal Statistics raster with the average mean canopy height of each study site as a raster cell value, as well as a Zonal Statistics table. A similar methodology was employed on only the DEM surface to calculate the average slope in percent rise for each study site with an output Zonal Statistics raster and table.

### Statistical Analysis Methods

A total of 7 LAI<sub>GROUND</sub>, 1 fCover<sub>GROUND</sub>, 6 LAI<sub>ALS</sub>, and 6 fCover<sub>ALS</sub> variables were calculated for statistical analysis. Seventeen other statistics were also derived from ALS data (Table 6).

There were 3 data traps: 10, 15, and 20-m, each calculated with (WOL) and without (NOOL) overlap for each dependent variable: LAI and fCover. Therefore, a 15-m LAI WOL variable is a 15-m data trap LAI variable calculated with overlap. All other ALS variables (percentiles and ranges, descriptive, ICVF, and environmental) were calculated using a 20-m data trap with overlap. Once all the variables from ground and ALS data were created, values were compiled and input into Stata 10 Statistical Modeling software. Stata 10 was used for all data modeling and statistical processing for this study. Five major statistical functions and tests were used in this study: single linear regression, principal component analysis (PCA), multivariate stepwise regression, Kmeans clustering analysis, and Pearson's Chi<sup>2</sup>.

Table 6: All ground and ALS derived variables used in this study.

<b>Study Variables</b>		
Ground		
<b>LAI</b>	Can-Eye effective Can-Eye true Miller effective	Miller true LAI2000(3) LAI2000(4) LAI2000(5)
<b>fCover</b>	Ground	
Airborne Laser Scanner (ALS)		
<b>LAI</b>	10 m NOOL 15 m NOOL 20 m NOOL	10 m WOL 15 m WOL 20 m WOL
<b>fCover</b>	10 m NOOL 15 m NOOL 20 m NOOL	10 m WOL 15 m WOL 20 m WOL
<b>Height Percentiles (m)</b>	5 25	50 75 95
<b>Percentile Ranges (m)</b>	Interquartile	5 - 95 Range
<b>Descriptive (m)</b>	Mean Standard Deviation	Skewness Kurtosis
<b>In Canopy Vegetation Fraction (ICVF)</b>	Low Vegetation	Medium Vegetation High Vegetation
<b>Environmental (m) (slope in percent rise)</b>	Average Elevation Mean Canopy Height	Average Slope

Regression was run with both LAI<sub>GROUND</sub> and fCover<sub>GROUND</sub> as dependent variables separately with LAI<sub>ALS</sub>, and fCover<sub>ALS</sub> values as independent variables. PCA was used to reduce ALS variables for input into stepwise regression to identify the most significant variables for LAI and fCover. The outcomes of the single and multivariate regression models were a function used to explain ground derived variables using ALS variables. Kmeans cluster analysis was used to identify 3 clusters of health status rank (HSr) from ground and LiDAR data, and was tested for agreement between themselves and with field survey estimated HSr values.

Single linear regression was performed twice, once with LAI and once with fCover. Prior to regression, a correlation matrix with two-tailed significance values was used to identify the most significant relationship between the ground LAI or fCover variable and the LAI<sub>ALS</sub> or fCover<sub>ALS</sub> variable. This process identified what data trap size performed best, and if overlap, or sampling rate, was significant in estimating LAI and fCover from the air. Regression analysis was performed, and the initial results used to identify if any outliers exists. A combination of visual analysis of the scatter plot, Studentized Residuals, and Cook's Distances were used as a means of identifying and explaining outliers. If they exist, outliers were removed, and a second regression was run.

Prior to principal component analysis, a correlation matrix with two-tailed significance values was created to identify independent variables that were highly correlated, or had similar correlation relationship with other variables, meaning that they shared common strong correlations. This matrix was used to remove redundant variables prior to PCA (Wold, Esbensen and Geladi 1987).

PCA is a statistical method used to extract the dominant patterns from a group of variables and detect the structure of the input variables. PCA has relatively few assumptions: (1) data have linearity in its relationships; (2) large variances have important structure; and (3) principal components are orthogonal, that they have no correlation between them. This method is used as a data reduction or exploration technique, taking a multitude of variables and reducing them into a small number of principal components (PCs). PCs are calculated to linearly capture the maximum variance of the original variables, while having no correlation between them. These PCs are ranked in order of the eigenvalue, or the amount of variability a single PC captures, and the sum of the eigenvalues is equal to the  $m$  number of variables used in PCA.

Each PC is also ranked on the proportion of total variation it captures, the eigenvalue divided by  $m$ , from 0.0 to 1.0. PCA will produce PCs until a user preset cut-off value, or until no trace of variability is left unexplained. Typically, however, only the top three to five eigenvalues are retained (Jolliffe 2005). While there are many methods for deciding how many PCs to keep, there are two easy to implement and standard methods: the Kaiser Criterion and the Scree Test (Zwick and Velicer 1986). The Kaiser Criterion states that PCs with eigenvalues greater than 1 should be kept. The assumption is that each original variable would have an eigenvalue equal to 1, so that any new component should have a value greater than 1 if it is to provide added benefit to analysis. The Scree Test looks at a plot of eigenvalue on the Y axis and PCs on the X axis. One should select the PCs where slope is the steepest; until the values level off on to create a graph that looks like a scree slope – a geologic term for the rocky debris that settles at the bottom of a steep slope. In most cases, these methods produce similar results. For this study, the Kaiser Criterion was used to determine the retained PCs.

While eigenvalues show how much variance a PC explains among all variables, a factor loading matrix shows how much of the PC variance is explained by the original input variables. Each component explains variance of some input variables, and the factor loading matrix displays which original input variables are most significant in explaining the variation in PCs. Through factor loading matrix analysis, one is able to draw conclusions about the relative significance of input variables in PCA.

The retained PCs from PCA are used as independent variables in a stepwise regression with ground LAI and fCover. Stepwise regression is a technique used to identify the most significant independent variables in a multivariate regression model. Stepwise regression can be forward or backward. Forward stepwise regression starts with an empty model with no

independent variables and adds independent variables one at a time. If the variable does not meet a user specified significance level when it is added to the regression, that variable is not retained. In backward stepwise regression, the analysis starts with a full model, or one with all the independent variables loaded. Each variable is then removed and its significance is tested. If it is determined that the removal of a variable is not significant, it is not retained. The result of both directions of stepwise analysis is a regression model that contains only significant independent variables. Ultimately, stepwise regression was used in this study to determine if a greater number of ALS derived statistics can better explain ground fCover and LAI better than single fCover<sub>ALS</sub> and LAI<sub>ALS</sub> variables. In this study, a threshold significance level of ( $p \leq 0.20$ ) was used in stepwise regression. This value allows for the inclusion of independent variables without being overly restrictive, especially with a small sample size.

The retained PCs from PCA were also used as input variables for Kmeans clustering analysis. This analysis was used to identify clusters in the ALS components data and the ground derived data. In total, 3 HSr clusters were created: those originally estimated in the field, those created from multivariable clustering analysis using ground data and those created from ALS principal component data. Using Pearson's Chi<sup>2</sup> Test, correlations were identified and tested to determine if significant clustering relationships existed between the ALS and ground derived clusters. In essence, this method was used to determine if it was possible to segment tree stands into rankings from ALS data, and tested if those rankings correlate significantly with ground rankings. Four tests were used to determine if the relationship between rankings were significant: Pearson's Chi<sup>2</sup> test, likelihood ratio Chi<sup>2</sup> test, Kendall's Tau-B test, and Fisher's exact test. By examining the values of these test, and the associated P values, a determination is made about whether it is possible to predict cluster ground health status from LiDAR data, and if those

clusters match what is predict from ground data. Though Pearson's  $\text{Chi}^2$  is not as trusted when expected cell values are less than 10, it is included and compared to the results of likelihood-ratio  $\text{Chi}^2$  for a better picture of significance between rankings.

## CHAPTER 4

### RESULTS

Results of this study include the single linear regression of ground and LiDAR derived measurement for LAI and fCover, Principal Components Analysis (PCA) of multiple LiDAR measurements, multivariate stepwise regression of ground data and LiDAR data for LAI and fCover, and Kmeans multivariate cluster analysis and comparison of those clustering results between ground and LiDAR data. Six sites originally surveyed and described in the study area, HCA # 71 – Dick’s Creek, are not included in the results due to the lack of LiDAR coverage.

#### Linear Regression Results

Single linear regression analysis was performed to determine how well airborne laser scanning (ALS) LAI and fCover proxy values could estimate those values derived from ground data. Correlations and two-tailed significance values were calculated for all combinations of variables, and the best performing pair was selected for examination of outliers and full regression analysis.

#### *LAI Regression Analysis*

Prior to calculation of coefficients of determinations, extreme and unrealistic values were removed from the dataset (where  $LAI_{ALS} > 15$ ), as well as those that could not be calculated due to division by zero. These cases occurred where no single or last returns were present in the vegetation class points, or where there were very few single and last returns above the 2-m cutoff that defined vegetation of interest. Accordingly, HCA-IDs 71-01, 71-02, and 71-04 were removed for 10 and 15-m traps, and 71-01 and 71-04 were removed for 20-m traps. In Table 7 and Figure

15,  $R^2$  coefficient of determination values for all ground LAI calculations and  $LAI_{ALS}$  values are shown.

Table 7:  $R^2$  coefficient of determination values for all combinations of ground and  $LAI_{ALS}$  values, with averages. Grey box values indicate significance at the 95% level. (\*) Denotes the highest  $R^2$ .

ALS Traps	Can-Eye		Miller		LAI2000			Avg.
	effective	true	effective	true	(3) Ring	(4) Ring	(5) Ring	
<b>No Overlap (NOOL)</b>								
<b>10m</b>	0.1643	0.2487	0.2512	0.2466	0.0116	0.1245	0.1011	<b>0.1640</b>
<b>15m</b>	0.1962	0.2863	0.2753	0.2189	0.0144	0.1269	0.1217	<b>0.1771</b>
<b>20m</b>	0.2704	0.3775	0.3076	0.3198	0.0978	0.2572	0.2693	<b>0.2714</b>
<b>Avg.</b>	<b>0.2103</b>	<b>0.3042</b>	<b>0.2780</b>	<b>0.2618</b>	<b>0.0413</b>	<b>0.1695</b>	<b>0.1640</b>	<b>0.2042</b>
<b>Overlap (WOL)</b>								
<b>10m</b>	0.2795	0.3891	0.3545	0.4151	0.1030	0.2488	0.2150	<b>0.2864</b>
<b>15m</b>	0.3479	0.4430	0.3586	0.3691	0.1025	0.2716	0.2595	<b>0.3074</b>
<b>20m</b>	0.3404	0.4565*	0.3452	0.3747	0.1735	0.3435	0.3607	<b>0.3420</b>
<b>Avg.</b>	<b>0.3226</b>	<b>0.4295</b>	<b>0.3527</b>	<b>0.3863</b>	<b>0.1263</b>	<b>0.2880</b>	<b>0.2784</b>	<b>0.3120</b>
<b>Avg.</b>	<b>0.2664</b>	<b>0.3669</b>	<b>0.3154</b>	<b>0.3240</b>	<b>0.0838</b>	<b>0.2288</b>	<b>0.2212</b>	<b>0.2581</b>

On average, all  $LAI_{ALS}$  proxy values estimated true ground LAI values (those calculated with an estimated clumping factor) better than effective values (those not calculated with an estimated clumping factor). Can-Eye true and Miller true LAI values had the strongest average coefficient values of all  $LAI_{ALS}$ , with 0.3669 and 0.3240, respectively. LAI2000 values averaged the worst overall with  $LAI_{ALS}$  proxy values. In almost all instances, larger data traps performed better than smaller data traps in estimating ground LAI.  $LAI_{ALS}$  proxy values calculated with overlap points estimated ground LAI values better than those sites that did not include overlap. Overall, larger data traps that include all available points generally provided the best coefficients

with true ground LAI measurements. The most significant correlation is between the 20-m data trap which included overlap to the Can-Eye true LAI with an  $R^2$  of 0.4564.

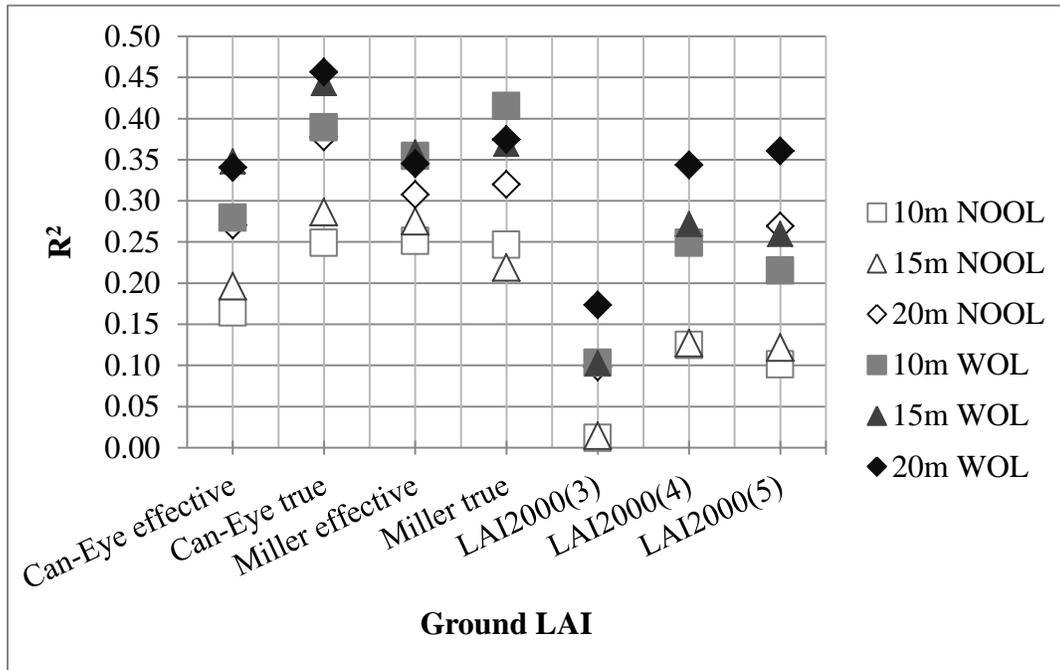


Figure 15:  $R^2$  values graphed for each  $LAI_{ALS}$  proxy matched with ground LAI values. Cases with LiDAR data overlap and no overlap are denoted as WOL and NOOL, respectively.

A full regression analysis was performed with the best performing ground and ALS coefficient, Can-Eye true LAI and 20-m Overlap LAI (Table 8). The regression produced an  $R^2$  of 0.4565, an Adjusted  $R^2$  of 0.4355, and an RMSE of 0.8123. To determine if any values represented outlier or anomalies in the data, the scatter plot, studentized residuals, and Cook's Distances were calculated and analyzed. In the LAI dataset it was determined that no significant outliers existed. The regression coefficient is 0.3456 with a standard error of 0.0740 and the constant coefficient is 3.0375 with a standard error of 0.2681. Both of these coefficients are

significant at the 95% level in the regression equation. Overall, 20-m LAI<sub>ALS</sub> with overlap can explain approximately 44% of the variation in Can-Eye true LAI.

Table 8: Regression analysis of Can-Eye True ground LAI and 20-m LAI<sub>ALS</sub> WOL. The analysis resulted with  $R^2 = 0.4565$ , adjusted  $R^2 = 0.4355$ , and RMSE = 0.8123.

Source	SS	df	MS			
Model	14.4074	1	14.4074	<b>Number of Obs</b>	28	
Residual	17.1563	26	0.6599	<b>F(1, 28)</b>	21.83	
Total	31.5637	27	1.1690	<b>Prop &gt; f</b>	0.0001	
				<b>R-squared</b>	0.4565	
				<b>Adj R-squared</b>	0.4355	
				<b>RMSE</b>	0.8123	
Can-Eye true	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
20m LAI <sub>ALS</sub>	0.3456	0.0740	4.67	0.000	0.1936	0.4977
Constant	3.0375	0.2681	11.33	0.000	2.4863	3.5886

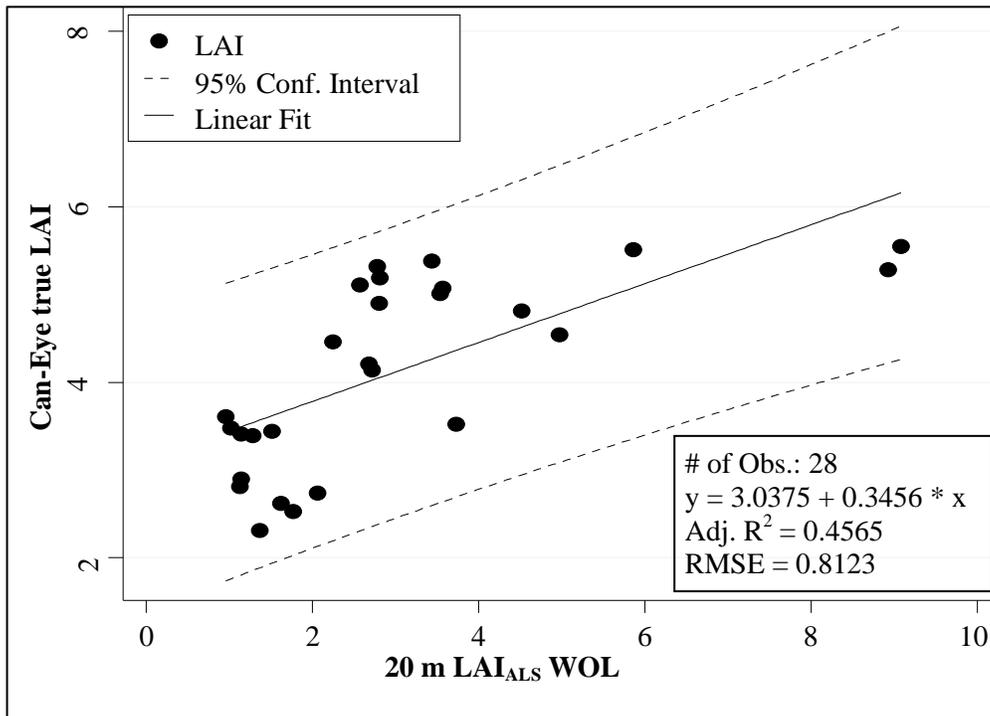


Figure 16: Regression of ALS derived LAI with the respective Can-Eye true value computer from hemispherical photography.

### *fCover Regression Analysis*

The same methodology was used to determine the goodness of fit between ground calculated fCover and ALS derived fCover. Again, all 6 data trap sizes WOL and NOOL were compared prior to selecting the size that performed the best. There is only a single fCover variable calculated from the Can-Eye software with which to compare the 6 ALS values.

The size of the data trap has the opposite result for fCover as it does with LAI. The 20-m traps, with and without overlap, perform the worst at calculating  $fCover_{ALS}$ , while the 15-m and 10-m traps do significantly better. With no overlap, the 10-m data trap performed best, and the 20-m data trap performed the worst. With overlap, the 15-m data trap performed best and the 20-m data trap performed the worst. However, the effect of overlap is still seen; all calculations using LiDAR data with overlap perform better than their no overlap counterparts of the same data trap size. Ground fCover was not initially as well predicted by  $fCover_{ALS}$  when compared to the results of LAI measurements. In the no overlap and overlap averages, as well as the overall averages, fCover coefficients were less than LAI coefficients. The best performing  $fCover_{ALS}$  estimator overall was the 15-m overlap, which was used for further regression analysis (Table 9 and Figure 17).

Table 9:  $R^2$  coefficients of determination values for all combinations ALS and ground fCover values, with averages. Grey box values indicate significance at the 95% confidence level. (\*)

Denotes the highest  $R^2$ .

ALS Traps	Ground fCover
<b>No Overlap (NOOL)</b>	
10m	0.1954
15m	0.1563
20m	0.0788
<b>Avg.</b>	<b>0.1435</b>
<b>Overlap (WOL)</b>	
10m	0.2240
15m	0.2290*
20m	0.0918
<b>Avg.</b>	<b>0.1816</b>
<b>overall Avg.</b>	<b>0.1625</b>

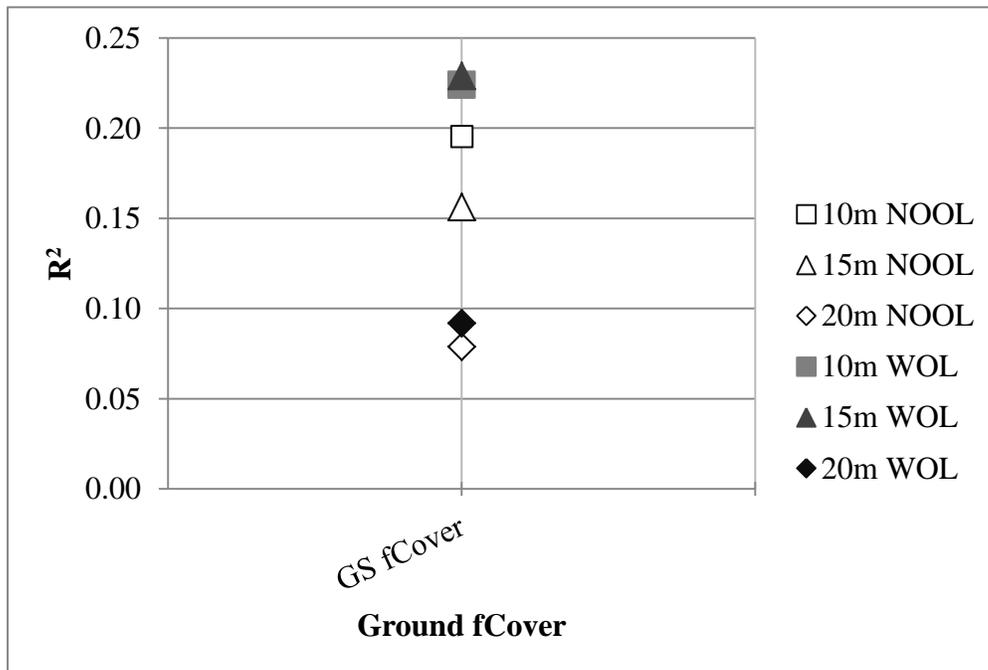


Figure 17: Graph of  $R^2$  values from all ALS data trap sizes compared to the single ground measured fCover.

Outlier analysis, using visual analysis of the scatter plot, studentized residual values and Cook's Distances, was performed to determine if site values should be removed from the final regression analysis (Table 10). Based on this analysis, three sites were suspected of being outliers: 71-01, 71-04, and 29-06. In general, studentized residuals approaching  $|3|$  are suspect. In this analysis, both 71-01 and 29-06 have studentized residuals approaching  $|3|$ , and each value is within the top 3 values of Cook's Distance. A third site, 71-04, with the fourth highest Cook's D and third highest absolute studentized residual, is also suspect based on these values and the fact that it shares the same scatter plot space as 71-01.

Table 10: Cook's Distance and Studentized Residuals for the highest and lowest five values of each, respectively. Grey box values indicate suspected outliers or anomalies.

<b>Rank</b>	<b>Cook's Distance</b>	<b>HCA-ID</b>	<b>Rank</b>	<b>Studentized Residual</b>	<b>HCA-ID</b>
<b>30</b>	3.93E-06	145-05	<b>30</b>	-2.992	29-06
<b>29</b>	0.0000179	29-09	<b>29</b>	-1.442	29-10
<b>28</b>	0.000099	29-05	<b>28</b>	-1.431	145-02
<b>27</b>	0.0001092	71-06	<b>27</b>	-1.327	29-01
<b>26</b>	0.0007823	145-04	<b>26</b>	-1.197	29-07
<b>5</b>	0.1053609	29-01	<b>5</b>	0.829	29-04
<b>4</b>	0.1575625	71-04	<b>4</b>	1.255	145-08
<b>3</b>	0.2242397	29-06	<b>3</b>	1.287	71-03
<b>2</b>	0.2248267	29-07	<b>2</b>	1.663	71-04
<b>1</b>	0.5875022	71-01	<b>1</b>	2.749	71-01

Data from the 3 outlier sites were removed from the final regression analysis because, on further inspection of site properties, and the vertical distribution of vegetation LiDAR returns, it is believed that canopy geometry and structure play a significant role in how these values are measured, and may explain why these values differ between the ground and LiDAR data. Both 71-01 and 71-04 have extremely high percentages of High Vegetation returns (greater than 50%

of all returns from the top 1/3 of the canopy) and very low percentages in the Middle and Low Vegetation returns (middle and lower 1/3 of canopy, respectively). Site 29-06 has a high percentage of Low Vegetation returns and very low percentages of Medium and High Vegetation returns and is the only site that has greater than 50% of all vegetation returns in the bottom 1/3 of its canopy. Sites 71-01 and 71-04 also were not included in LAI regression because they had very few single and last canopy returns yielding LAI values of greater than 100 – further confirming the decision that these sites may represent anomalies in the dataset. The final regression analysis, with 3 outlier sites removed, can be seen in Table 11 and Figure 18.

The goodness of fit was greatly improved over the original results, with an  $R^2$  of .4805, adjusted  $R^2$  of .4597, and an RMSE of .1581. The coefficient of determination was almost doubled with the removal of three sites. RMSE was slightly improved, with only a difference of 0.0613 between the original analysis and the final one with outliers removed. The variable coefficient is 1.2105, and was found significant in the analysis; however the constant coefficient was -0.1986 and found to be not significantly different from zero. The 95% confidence also was slightly improved and narrowed from the original regression analysis. Overall,  $fCover_{ALS}$  alone can explain approximately 46% of the variation found in ground  $fCover$ .

Table 11: Regression analysis of ground fCover and 15m fCover<sub>ALS</sub>.WOL. The analysis resulted with  $R^2 = 0.4805$ , adjusted  $R^2 = 0.4597$ , and  $RMSE = 0.4597$ .

Err. Source	SS	df	MS		
Model	0.5781	1	0.5781	Number of Obs	27
Residual	0.6250	25	0.0250	F(1, 28)	23.13
Total	1.2031	26	0.0463	Prop > f	0.0001
				R-squared	0.4805
				Adj R-squared	0.4597
				RMSE	0.1581
ground fCover	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
15m fCover <sub>ALS</sub>	1.2105	0.2517	4.81	0.000	0.6921 1.7290
Constant	-0.1986	0.1878	-1.06	0.300	-0.5853 0.1881

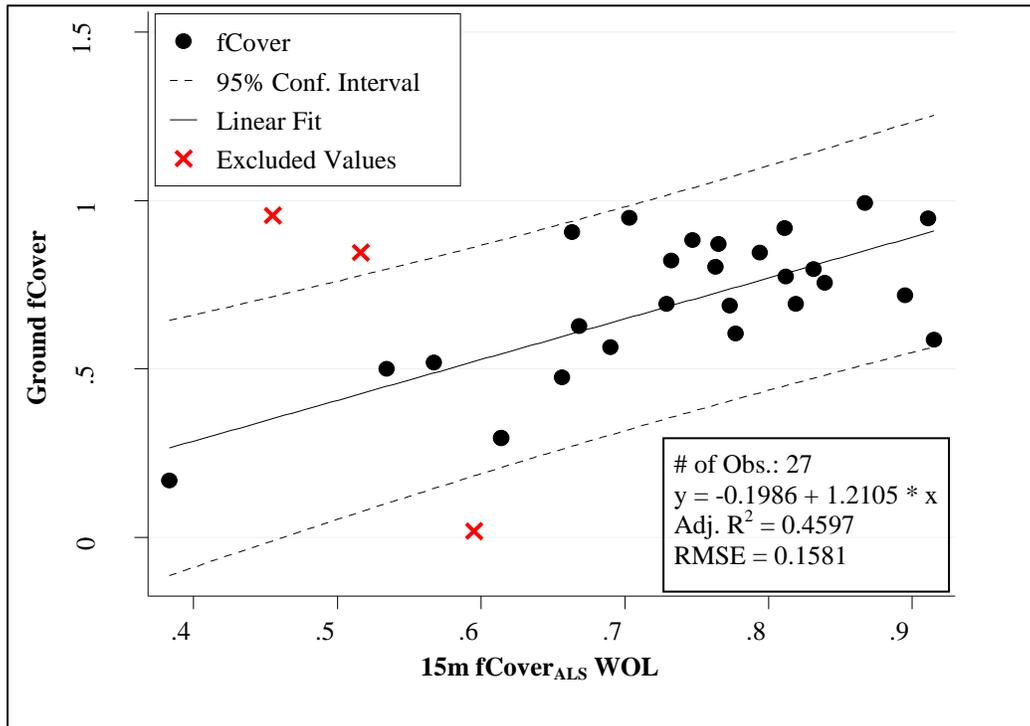


Figure 18: Regression of ALS derived fCover with the respective value computer from hemispherical photography.

### PCA and Stepwise Multivariate Regression Results

Prior to performing PCA, all variables underwent correlation analysis, and redundant variables were removed. In total, 5 variables were removed: 5 percentile height, 95 percentile height, 5 – 95 percentile height range, and mean height. Elevation also was excluded because it did not provide additional prediction power in PCA. Each site had exclusive elevation ranges, and because of this, elevation caused artificial clustering of values. It was determined that elevation should be excluded because ultimately said more about spatial location than it did about how HWA affects hemlock defoliation. The remaining 12 variables, plus the best performing  $LAI_{ALS}$  and  $fCover_{ALS}$  metrics determined from single linear regression were used as original input variables, for a total of 14 ALS derived variables used in PCA. Sites 71-01 and 71-04 were excluded from PCA because they lacked  $LAI_{ALS}$ , and are suspect for being anomalies in the dataset. The number of principal components was not capped for PCA, and the reported values were not rotated (Table 12).

Table 12: Results from principal component analysis shows a total of 14 components produced. Following the Kaiser Criterion, only those components with eigenvalues great than 1 are retained, shown in grey boxes. These 5 components explain approximately 91% of all variance from the original input variables.

<b>Component</b>	<b>Eigenvalue</b>	<b>Difference</b>	<b>Proportion</b>	<b>Cumulative</b>
<b>Comp1</b>	6.0689	3.4444	0.4335	0.4335
<b>Comp2</b>	2.6246	0.6521	0.1875	0.6210
<b>Comp3</b>	1.9724	0.8638	0.1409	0.7618
<b>Comp4</b>	1.1086	0.1009	0.0792	0.8410
<b>Comp5</b>	1.0077	0.4578	0.0720	0.9130
<b>Comp6</b>	0.5499	0.2102	0.0393	0.9523
<b>Comp7</b>	0.3396	0.1634	0.0243	0.9765
<b>Comp8</b>	0.1763	0.1042	0.0126	0.9891
<b>Comp9</b>	0.0721	0.0177	0.0052	0.9943
<b>Comp10</b>	0.0544	0.0375	0.0039	0.9982
<b>Comp11</b>	0.0169	0.0085	0.0012	0.9994
<b>Comp12</b>	0.0084	0.0081	0.0006	1.0000
<b>Comp13</b>	0.0003	0.0003	0.0000	1.0000
<b>Comp14</b>	0.0000	0.0000	0.0000	1.0000

The eigenvalue, or total variance accounted for by each component, is listed from highest to lowest. The individual and cumulative proportions also are listed along with the differences between each eigenvalue. All of these values describe how well a single component captures the variability of the input data. Following the Kaiser Criterion, only those components with an eigenvalue of greater than 1 are retained – for a total of 5 PCs. Together, these 5 PCs explain approximately 91% of the total variance from all the original ALS input variables. Notice that while 14 components were produced, the eigenvalue and proportion of variance explained drops off significantly towards the bottom, and component 14 reports only a trace of explained variance (though the eigenvalue is 0.000, a trace amount is still present).

Table 13: Factor Loading Matrix from principal component analysis. Grey box values represent loadings greater than  $\pm 0.3000$ .

Variable	Comp1	Comp2	Comp3	Comp4	Comp5
<b>20m Ave Slope</b>	0.0931	0.2766	0.3918	0.1407	-0.3016
<b>Mean Can H</b>	0.3866	0.0963	0.1169	0.0819	0.0238
<b>15 m FC WOL</b>	0.0835	0.1444	0.4823	-0.4652	0.2064
<b>20 m LAI WOL</b>	0.1156	-0.2263	-0.3925	0.3640	-0.3961
<b>25 percentile</b>	0.3454	-0.1991	0.2150	0.2096	0.0355
<b>50 percentile</b>	0.3901	0.0167	0.0551	0.0963	0.0136
<b>75 percentile</b>	0.3826	0.1413	0.0323	0.1618	0.0056
<b>IQ Range</b>	0.1927	0.4925	-0.2236	-0.0003	-0.0365
<b>Std. Dev.</b>	0.1734	0.5329	-0.0965	0.1725	0.0035
<b>Skewness</b>	-0.3142	0.2004	0.1128	0.3242	0.0821
<b>Kurtosis</b>	-0.0865	-0.1522	0.2652	0.6034	0.5567
<b>Low Veg</b>	-0.3008	0.3567	-0.1704	0.0809	0.1687
<b>Med Veg</b>	-0.1551	-0.0767	0.4526	0.1321	-0.5628
<b>High Veg</b>	0.3401	-0.2434	-0.1399	-0.1495	0.2105

Factor analysis from factor loading matrices can be used to interpret the influence of each independent variable on the creation of PCs. Since only 5 PCs were retained after the initial analysis, only those are shown in the factor loading matrices (Table 13). The factor loading matrix shows that the Mean Canopy Height, 25, 50 and 75 percentile, skewness, and the Low and High canopy fractions all load approximately equally in the first component (PC1). No one variable shows significant loading over the others. This suggests that these variables share similar variance space. The second PC (PC2) has stronger factor loadings than the first one, with IQ Range, Standard Deviation of Height AGL, and Low Vegetation fraction greater than 0.3 loading. In PC 3 and 4, our most significant  $LAI_{ALS}$  and  $fCover_{ALS}$  variables from single linear regression both load strongly.

*LAI Regression Analysis*

All 5 retained PCs were used as independent variables in stepwise regression, with Can-Eye true LAI as the dependent variable. Significance level was set at ( $p \leq 0.20$ ) for both forward and backward regression. Both regression analyses results were in agreement with the significant PCs of the regression, so only one regression result representing both directions is shown (Table 14).

Table 14: Forward and backward stepwise regression results were identical in results for multivariate LAI. The regression produced an  $R^2 = 0.7703$ , and Adj.  $R^2 = 0.7303$ , and an RMSE = 0.5615.

Source	SS	df	MS		Number of Obs	28
Model	24.3132	4	6.0783		<b>F(1, 28)</b>	19.28
Residual	7.2505	23	0.3152		<b>Prop &gt; f</b>	0.0000
Total	31.5637	27	1.1690		<b>R-squared</b>	0.7703
					<b>Adj R-squared</b>	0.7303
					<b>RMSE</b>	0.5615
Can-Eye true	Coef.	Std. Err.	t	P> t		Beta
PC1	0.2572	0.0439	5.86	0.000		-0.5908
PC2	-0.3943	0.0667	-5.91	0.000		0.5859
PC3	-0.1640	0.0769	-2.13	0.044		-0.2130
PC5	-0.1944	0.1076	-1.81	0.084		-0.1805
Constant	4.0646	0.1061	38.31	0.000		

This regression retained four of the five original PCs, and reports an  $R^2 = 0.7703$ , an Adjusted  $R^2 = 0.7303$ , and an RMSE = 0.5615. The multivariate solution, therefore, shows a great improvement in estimating ground LAI over single linear regression (Adj.  $R^2 = 0.4355$ ) and explains approximately 73% of the variability found in ground LAI measurement. The results show that although a relatively high significance level for elimination was given, all variables are

significant at the ( $p \leq 0.10$ ) level, with two at the ( $p \leq 0.05$ ) level. In multivariate regression, a standardized coefficient, or Beta, can be calculated to determine the level of effect an independent variable has on the dependent variable. In this result, we see that PC1 has the most effect on the regression result with a Beta of -0.5908, followed closely by PC2. This model kept PC3 and PC5, two PCs where the original LAI<sub>ALS</sub> variable loaded highly, but did not keep PC4, which also had a high loading factor from LAI<sub>ALS</sub>.

### *fCover Regression Analysis*

The same methodology was used for multivariate stepwise regression for fCover. Again, all 5 retained PCs were used as independent variables and ground derived fCover was the dependent variables. Both forward and backward stepwise regressions were in agreement about the most significant variables and the results were identical (Table 15).

Table 15: Forward and backward stepwise regression results were identical in fCover Results.

The regression produced an  $R^2 = 0.5456$ , an Adj.  $R^2 = 0.4666$ , and an RMSE = 0.1801.

Source	SS	df	MS		Number of Obs	28
Model	0.8956	4	0.2239		<b>F(1, 28)</b>	6.90
Residual	0.7460	23	0.0324		<b>Prop &gt; f</b>	0.0008
Total	1.6416	27	0.0608		<b>R-squared</b>	0.5456
					<b>Adj R-squared</b>	0.4666
					<b>RMSE</b>	0.1801
Ground fCover	Coef.	Std. Err.	t	P> t		Beta
pc4	-0.1128	0.0329	-3.43	0.002		-0.4818
pc3	0.0811	0.0247	3.29	0.003		0.4619
pc1	0.0253	0.0141	1.80	0.085		0.2527
pc2	-0.0290	0.0214	-1.35	0.189		-0.1902
Constant	0.6683	0.0340	19.63	0.000		

Multivariate stepwise regression analysis using principal components showed little improvement over the single linear regression results ( $\text{Adj. } R^2 = 0.4597$ ). The PCA  $R^2 = 0.5456$  is improved, but when adjusted for the population in Adjusted  $R^2$ , the model shows minimal improvement. As with single linear regression, fCover is not as well predicted as LAI. In this model, four PCs were found significant at the ( $p \leq 0.20$ ) level, and two of those PCs were significant at the ( $p \leq 0.05$ ) level. These highly significant PCs showed strong influence on the dependent variables with Betas of greater than  $|0.45|$ . The stepwise selection kept PC3 and PC4 as the two most significant variables in the regression, and the fCover<sub>ALS</sub> variable had high loadings in both components.

#### Multivariate Health Status Clustering Results

The final analysis, using Kmeans clustering, is designed to derive clusters from ALS PCs and from ground LAI and fCover, and to test to see if those clusters are significantly related to each other—these results aim to show that it is possible to assign Health Status ranking (HSr) to a field site with reasonable confidence from ALS data which agrees with the HSr that would be derived from ground LAI and fCover data. These two sets of clusters also were compared to those HSr assigned during field survey.

For ALS cluster analysis, all 5 retained PCs with eigenvalues greater than 1 were used as variables to derive the clusters, labeled “ALS PC HSr”. For ground cluster analysis, fCover and Can-Eye true LAI area used, labeled “GS HSr”. A cluster value of  $k = 3$  was used for each cluster analysis to align with the 3 value rank used in field survey, labeled “Field HSr”. Each table is listed with the related significance tests (Table 16, Table 17, and Table 18). Each cell contains three values: the actual frequency, the expected cell frequency, and the  $\text{Chi}^2$  contribution value, used in the calculation of Pearson’s  $\text{Chi}^2$ .

Neither the ALS PC HSr nor GS HSr was found to be significantly related to Field HSr by any of the four tests used to determine relationship and significance (Table 16). The comparison of Field HSr and GS HSr resulted in a Pearson's  $\chi^2 = 2.5769$  and a Pr = 0.631, likelihood-ratio  $\chi^2 = 3.4818$  and a Pr = 0.481, Kendall's Tau-B = 0.0455, and a Fisher's exact value = .719. By no means or tests are the rankings of these two variables significantly related to field-based measurements.

Table 16: Comparison of rankings between Field HSr and GS HSr. This relationship is not significant by any test: Fisher's exact, likelihood-ratio  $\chi^2$ , or Pearson  $\chi^2$ .

		Field HSr					
GS HSr		1	2	3	Total		
1	3	7	3	13	Freq.		
	3.3	6.5	3.3			13	Exp. Freq.
	0.0	0.0	0.0			0.1	Chi2 Cont.
2	4	4	3	11			
	2.8	5.5	2.8			11	
	0.6	0.4	0.0			1	
3	0	3	1	4			
	1.0	2.0	1.0			4	
	1.0	0.5	0.0			1.5	
Total	7	14	7	28			
	7.0	14.0	7.0			28	
	1.6	0.9	0.0			2.6	
Pearson $\chi^2(4) =$		2.5769	Pr =	0.631			
likelihood-ratio $\chi^2(4) =$		3.4818	Pr =	0.481			
Kendall's tau-b =		0.0455	ASE =	0.150			
Fisher's exact =		0.719					

The results with Field HSr and ALS PC HSr are similar (Table 17). This comparison resulted in a Pearson's  $\chi^2 = 2.4381$  and a  $Pr = 0.656$ , likelihood-ratio  $\chi^2 = 2.4193$  and a  $Pr = 0.659$ , Kendall's Tau-B =  $-0.0664$ , and a Fisher's exact value =  $.706$ . By no means or tests are the rankings of these two variables significant. In both comparisons, although the Pearson's  $\chi^2$  is not expected to perform as well with cell counts below 10, the values are nearly identical to the likelihood-ratio  $\chi^2$  values.

In both contingency tables, there does not appear to be a systematic difference in the variables estimated with ALS PC HSr and GS HSr when compared to Field HSr. In many cases, however, the difference was a matter of a single rank, either between 1 and 2, or 2 and 3. In a much smaller percentage of cases values were off by 2 ranks, indicating a large discrepancy between what was marked in the field and what was calculated by Kmeans clustering. Between Field HSr and GS HSr, only 3 sites were off by more than 1 ranking, and in Field HSr and ALS PC HSr, only 4 sites were off by more than 1 ranking.

Table 17: Comparison of rankings between Field HSr and ALS PC HSr. This relationship is not significant by any test: Fisher's exact, likelihood-ratio  $\chi^2$ , or Pearson  $\chi^2$ .

ALS PC HSr	Field HSr			Total	
	1	2	3		
<b>1</b>	2 1.8 0.0	2 3.5 0.6	3 1.8 0.9	<b>7</b> <b>7</b> <b>1.6</b>	Freq. Exp. Freq. Chi2 Cont.
<b>2</b>	4 3.8 0.0	8 7.5 0.0	3 3.8 0.1	<b>15</b> <b>15</b> <b>0.2</b>	
<b>3</b>	1 1.5 0.2	4 3.0 0.3	1 1.5 0.2	<b>6</b> <b>6</b> <b>0.7</b>	
<b>Total</b>	<b>7</b> <b>7.0</b> <b>0.2</b>	<b>14</b> <b>14.0</b> <b>1.0</b>	<b>7</b> <b>7.0</b> <b>1.2</b>	<b>28</b> <b>28</b> <b>2.4</b>	
<b>Pearson chi2(4) =</b>		2.4381	<b>Pr =</b>	0.656	
<b>likelihood-ratio chi2(4) =</b>		2.4193	<b>Pr =</b>	0.659	
<b>Kendall's tau-b =</b>		-0.0664	<b>ASE =</b>	0.179	
<b>Fisher's exact =</b>		0.706			

When comparing the ALS PC HSr and the GS HSr, a significant distribution between the two ranking variables is apparent (Table 18). Both the Pearson's  $\chi^2$  Pr and likelihood-ratio  $\chi^2$  Pr= 0.000, and Fisher's exact value = 0.000. Kendall's Tau-B, which measures both the strength and direction of the relationship between the two rankings have a value of 0.4664, indicating a strong relationship that is moderately position in nature. Of the total 28 observations, the ALS PC HSr clusters and GS HSr clusters agree on 21 of those, or 75%.

Table 18: Comparison of rankings between GS HSr and ALS PC HSr. By all measures, this relationship is significant.

ALS PC HSr		GS HSr			Total	
		1	2	3		
<b>1</b>	7	0	0	<b>7</b>	Freq.	
	3.3	2.8	1.0	<b>7</b>		Exp. Freq.
	4.3	2.8	1.0	<b>8.1</b>		Chi2 Cont.
<b>2</b>	3	11	1	<b>15</b>		
	7.0	5.9	2.1	<b>15</b>		
	2.3	4.4	0.6	<b>7.3</b>		
<b>3</b>	3	0	3	<b>6</b>		
	2.8	2.4	0.9	<b>6</b>		
	0.0	2.4	5.4	<b>7.7</b>		
<b>Total</b>	<b>13</b>	<b>11</b>	<b>4</b>	<b>28</b>		
	<b>13.0</b>	<b>11.0</b>	<b>4.0</b>	<b>28</b>		
	<b>6.6</b>	<b>9.5</b>	<b>7.0</b>	<b>23.1</b>		
<b>Pearson chi2(4) =</b>		23.1000	<b>Pr =</b>	0.000		
<b>likelihood-ratio chi2(4) =</b>		25.8568	<b>Pr =</b>	0.000		
<b>Kendall's tau-b =</b>		0.4664	<b>ASE =</b>	0.186		
<b>Fisher's exact =</b>		0.000				

There were two noticeable disagreements in the contingency table where GS HSr rank was 1 and ALS PC HSr was ranked as 2 or 3. GS HSr ranked 13 sites as HSr 1, while ALS PC HSr only ranked 7 sites as HSr 1. The additional 6 sites were split evenly between HSr 2 and 3 from ALS PC HSr. Of the 7 sites that were not in agreement, 4 of those are off by a single rank and 3 are off by 2 ranks. This may indicate that instead of a single threshold value that determines the differences between ranks, there is a buffer or grey area that exists. Since all these variables are continuous, it seems logical that at least some site rankings will fall in between “true” discrete rankings of 1, 2, and 3. However, with an agreement of 75% of the total sites, and

significance indicated by all measures, this test shows that it is possible to estimate tree stand health status in a 3 class ranking from ALS data.

To gain better insight into why 7 sites did not agree between Ground clustering and ALS clustering, the sites were inspected for any additional factor that may lead to misclassification of health status by Ground or ALS clustering. No pattern was detected in the data that would indicate that these 7 sites have something in common that leads to the discrepancy in rank. For comparison purposes, Table 19 lists the sites and their associated ALS and GS HSr values.

Table 19: HCA IDs and their associated rankings from Ground and ALS clustering. Sites 71-01 and 71-04 were excluded from clustering analysis because they lacked LAI values. Grey box values indicate disagreement between GS and ALS PC HSr.

HCA ID	ALS PC HSr	GS HSr	HCA ID	ALS PC HSr	GS HSr
145-01	2	1	29-08	2	1
145-02	3	3	29-09	2	1
145-03	2	3	29-10	1	1
145-04	2	2	29-11	1	1
145-05	3	3	29-12	3	1
145-06	3	3	71-01	---	---
145-07	2	2	71-02	2	2
145-08	2	2	71-03	2	2
29-01	3	1	71-04	---	---
29-02	3	1	71-05	2	2
29-03	1	1	71-06	2	2
29-04	1	1	71-07	2	2
29-05	1	1	71-08	2	2
29-06	1	1	71-09	2	2
29-07	1	1	71-10	2	2

## CHAPTER 5

### DISCUSSION

#### Environmental Considerations

Although this study aims to determine the effectiveness of ranking the health status of hemlock tree stands in north Georgia, not a wide range of infestation was found in the field. Within the hemlock stands surveyed, there is a fairly homogeneous spread of HWA infestation. Although differences exist in the duration of infestation, on a gradient from east to west and with east representing longer infestation times, no significant directional bias was found. HWA was present in all areas and sites surveyed. That is to say, no one area appears to be more infested, or in more dire health than another. However, specific sites do show minor health variation. Because the range of health was narrow, the tolerances of field-determined health clusters became more arbitrary. Health, as measured in this study, is a continuous entity based on LAI and fCover. Many sites did not fall exactly into a single health ranking, but rather straddled the edge between two. However, as the GS and ALS HSr clustering results show, clusters can be identified both from the ground and LiDAR data, and those clusters generally agree. One major environmental variable that showed significant influence on cluster values was elevation. This value was removed from the clustering analysis because it hid the variation of other variables. Since each HCA surveyed had a unique elevation range, and elevation ranges did not overlap between areas, including elevation as a variable tended to artificial cluster sites not based on the health of the trees found there but the spatial location and elevation from which it was surveyed. However, LAI and fCover values from the ground did not indicate spatial significance or

elevation significance; therefore, it was determined to not explain how HWA affects tree defoliation.

Each site was required to have at least fifty percent minimum hemlock species. In general, these sites contained greater percentages of hemlock and were mixed with other needle-leaf species. However, lack of “pure” hemlock stands made it difficult to specifically say what the underlying cause of differences were in values seen in this study: specific environmental factors, or the influence of HWA. However, a strong attempt was made to keep environmental characteristics between all sites the same and the results from this study still show significant differences in values and clusters from both ground and ALS data.

Some error will result from the differences in collection of LiDAR data and field data because almost one year passed between the collections of these data. USFS monitoring of HWA spread in Georgia has showed slow progress of the infestation throughout the state, indicating that the results from this study can still be considered valid despite the temporal difference. Some error between ground and ALS data exists due to the collection date, but it should not be considered a significant factor. Lastly, some error will come from the collection and processing of both the LiDAR and ground data, both due to operator and random error. This is to be expected. However, despite these contributions of error and uncertainty in the data, the results still show that a relationship exists, and that HWA influence can be measured remotely.

#### Data Equation Considerations

The equation used for  $LAI_{ALS}$  may not be the most appropriate because it tends to encounter division by zero errors or extremely overestimate values ( $LAI > 100$ ) where there is a deficit of last and single returns from the canopy. This is because it is calculated as a fraction of first returns to last and single returns. A better equation or derived LAI proxy less sensitive to the

relationship between first, last, and single returns may have provided better results. The inclusion of intermediate returns in an equation derivation may provide better results and take advantage of the multiple return sensors types almost exclusively used now for LiDAR collection. In other studies, a simple transformation of canopy to ground return percentages has been used to approximate LAI, but this type of equation can be sensitive to the type of sensor used and the canopy geometry.

In previous studies, LiDAR LAI generally best estimated effective ground LAI values; however, in this study it was found that these values best correlated with true estimations of LAI, where a clumping factor was applied. In general, it is likely that  $LAI_{ALS}$  will underestimate the true LAI as measured from the ground, which is why it correlated so well with effective measures. A proposed hypothesis is that the  $LAI_{ALS}$  calculation used in in this study systematically over estimates effective LAI when compared to previous equations used. This may be of added benefit to those interested in calculating LAI from ALS, because in general, true LAI is the desired measure for most forest studies.

The calculation of fCover in this study is a widely used and accepted equation, although it too has pitfalls. Though fCover from the ground is calculated as the percent vegetation covering ground for a given area, with ALS there is generally no way to know what percentage of the pulse footprint is constituted by vegetation and ground from a multiple return sensor. When intermediate returns are filtered from the data prior to calculation, it adds an additional uncertainty in truly understanding what percentage of the ground was exposed during LiDAR collection. For example, a pulse travels towards a canopy and returns four times – three of those times are vegetation returns. What percentage of the laser footprint was actually exposed ground in the pulse footprint? Using discrete return data, there is no true way to know, and therefore,

fCover<sub>ALS</sub> is a best estimation. The inclusion of an actual percentage of pulse returned by ground or vegetation may give more insight to a better fCover calculation.

### Canopy Geometry in ALS Measurements

fCover and LAI measurements derived from ALS data are very sensitive to canopy geometry. As seen with sites 71-01, 71-04, and 29-06, variable canopy structure, especially canopies with highly uneven canopy point return distributions, or high values in either the low or high canopy, are not measured as well. fCover and LAI from the ground are measured in 2D space, capturing only a flat image of the canopy cover from below and calculating values from that space. From ALS data, these variables are measured in 3D space, considering the vertical distribution of all canopy elements. Therefore, when a high percentage of points are returned from only a single level in the canopy, LiDAR will tend to miscalculate these variables when compared to hemispherical ground measurement. When dealing with only first, last, and single return data, a negatively skewed distribution with a slightly positive kurtosis is expected, because the majority of first and single returns will come from towards the top of the canopy. On average, skewness for all sites was -0.2033 and kurtosis was 2.6095. However, when this distribution is altered, either to be extremely negatively skewed as with 71-01 and 71-04, or positively skewed, as is the case with 29-06, fCover and LAI tend to be poorly predicted. In fact, 71-01 has the most negative skewness of all study sites (-1.07), while 29-06 has the most positive skewness of all sites (1.22). In addition, 29-06 and 71-01 rank first and fourth in kurtosis.

### Data Trap Size and Sampling Rate

The ALS derived values of fCover and LAI are sensitive to both the data trap size used to select the LiDAR points and the sampling rate, or total pulses per site, to calculate variables. The results of this study suggest that different data trap sizes should be considered for calculation of

LAI and fCover, and that the optimal data trap size for LAI may not be the optimal size for fCover calculation. From the ground using hemispherical photography, LAI and fCover values are calculated from different angles. In this study, a 60° Circle of Interest was used to calculate LAI while fCover was calculated from the integration of angles 0 - 10°. In this study, 20-m data traps performed best with LAI, with 15-m data traps performing second best on average. The 15-m data traps performed best with fCover, although 10-m data traps were a close second and even outperformed the 15-m trap without overlap. As might be expected, smaller data traps tended to do better with fCover, while larger data traps tended to do better with LAI. Previous studies have tested the significance of data trap sizes on coefficients of determination with hemispherical photography, but these studies generally settle on a single data trap size for calculating both values. This study suggests that different optimal sizes should be considered for all metrics, and that it cannot be assumed that a single data trap size is optimal for all ALS derived forest metrics.

The inclusion or exclusion of overlap points in calculating forest metrics from ALS data served as a proxy to determine if LiDAR sampling rate played a role in how well metrics could be estimated from the air. In this study, higher LiDAR sampling rates of canopy led to better estimations of forest metrics, as shown in the improved performance of sites calculated with overlap versus those calculated without overlap. In all cases in this study for LAI and fCover, data traps with overlap outperformed data traps of the same size without overlap, all other things being equal. The overlap versus no overlap comparisons show that a higher sampling rate, or higher post spacing in the collection of LiDAR, may lead to better calculations of LAI and fCover. Although the statistical significance of this improvement was not considered, it still shows promise in more accurately calculating forest metrics. Increasing LiDAR sampling can be achieved either through collection methods (e.g., flying lower and slower), higher sensor

collection frequencies, or making multiple flying passes over a single area – as was simulated in this study.

### Prediction Power of Additional ALS Variables

Mixed results were achieved when additional ALS derived variables were used in an attempt to more accurately predict ground variables. LAI results saw significant improvements in prediction ability with the inclusion of additional variables through principal component analysis (PCA). Height Percentiles and Ranges, Mean Canopy Height and Low and High Vegetation Fractions loaded highly in the first two retained principal components, which explained approximately 62% of the variability (PCs), while the original ALS derived fCover and LAI variables, as well as skewness, kurtosis, Medium Vegetation Fraction and slope loaded highly in the last three retained PCs, for a total explained variability of approximately 91%. In general, in an independent-dependent variable correlation matrix, many variables tended to correlate well with LAI prior to PCA, especially In Canopy Vegetation Fractions. In fact, some of these additional variables correlated better than the  $LAI_{ALS}$  proxy value used in single linear regression. Low Vegetation Canopy Fraction and High Vegetation Canopy Fraction were strongly correlated with Can-Eye true LAI ( $R^2 = 0.5847$  and  $R^2 = 0.4991$ , respectively, in single linear regression). It makes sense that PCA would produce ALS PCs that better predict ground LAI than the  $LAI_{ALS}$  proxy alone when combined with other correlated values such as Height Percentiles.

With fCover, however, PCA did not provide much added prediction power. In an independent-dependent variable correlation matrix, only the original  $fCover_{ALS}$ , skewness, and Low Vegetation Fraction were correlated with ground fCover, and no correlations were stronger than the original single linear regression with only fCover variables. Therefore, in this study,

fCover was poorly predicted overall, even with the addition of other LiDAR variables. No patterns in the data suggest why fCover is so poorly correlated with additional LiDAR metrics.

#### Agreement among ALS, Ground and Field HSr

Field health rankings (HSr) were not found to be significantly related to either ground or LiDAR based health rankings. Field rankings are subject to the interpretation of the surveyor, the current site conditions, and a host of other factors. Field survey also may not be able to distinguish as well between sites of similar conditions as quantitative mathematical measurements. Although this study shows a weak agreement between qualitative field assessment and quantitative ground and ALS measurement, this does not suggest that one is more correct. It does indicate that considering both qualitative and quantitative methods when determining the health status of forest stands may lead to better indications of tree health.

Kmeans clustering analysis in this study demonstrated that three significantly different clusters could be extracted from both ALS and ground data and those clusters generally agreed. This is perhaps the most significant result from the entire study. The results show that it is possible to partition values, and that with 75% agreement, ranking of forest health from ALS data matches that of ground data. In its current state, the process of clustering is a complex multistep statistical process which requires the computation of many LiDAR variables. However, using this process, the results of the study suggest that it is possible to estimate a tree stand health ranking using only LiDAR data that would agree with ground data assessment approximately 75% of the time.

## CHAPTER 6

### CONCLUSION

Overall, the objectives of this study were met. However, each objective of this study was accomplished with varying degrees of success. Correlation analysis was used to determine the LiDAR derived LAI and fCover value which best matched ground measures of LAI fCover. These values were used in single linear regression to determine the goodness of fit, which resulted in moderately good correspondence between air and ground data, with Adj.  $R^2 = 0.4355$  and  $0.4597$  for LAI and fCover, respectively. Single linear regression showed sensitivity to canopy geometry, and indicated better equations for LAI and fCover from LiDAR data may be needed to achieve more reliable results. In most cases, LAI values were in greater agreement than fCover values, though no strong hypothesis exists for why this is so. Additional variables were added and processed using principal component analysis and stepwise regression to determine if the addition of other LiDAR variables improved in the estimation of ground LAI and fCover. This analysis had mixed results, with LAI being significantly improved with the additional variables, and fCover showing little to no improvement, with Adj  $R^2 = 0.7307$  and  $0.4666$  for LAI and fCover, respectively. Lastly, Kmeans cluster analysis was used to derive 3 clusters of sites from ground and LiDAR principal components data. Clustering analysis proved to be the most meaningful and interesting result from this study. Approximately 75% of LiDAR and ground derived clusters were in agreement after analysis, and this correlation was found significant by all measures. This result shows the health of a hemlock tree stand can be ranked from LiDAR data with a fairly high level of certainty, and those rank clusters have a high level

agreement with ground health ranking. This result, above all else, shows the objectives of this study were met, i.e., clustering of tree stand health is possible.

Additionally, this study finds sampling rate and data trap size play an important role in how well LiDAR measures predict ground measures. The goodness of fit varied by data trap size for LAI and fCover, and the best performing sizes were not the same for each variable. Larger data trap sizes were found to better predict LAI (20 m), while small sizes better predicted fCover (10 or 15 m). In almost all cases, increasing the sampling rate improved the goodness of fit for all LiDAR variables. For LAI measures, the equation used in this study better predicted true measures of LAI better than effective measures.

One of the greatest challenges of this study was locating appropriate field sites for study with available LiDAR data. Being on the southern end of hemlocks' natural range, and heavily mixed with pine, it was difficult to locate appropriate stands. Often, hemlocks in Georgia are limited to cool, steep slopes near rivers and streams. Where they exist outside of these habitats, they are generally a secondary species. Although remote sensing techniques for finding HWA infested hemlock stands exist, ultimately field site selection in this study was achieved the old-fashioned way—by following hunches and exploring in the woods.

In future studies, more attention should be given to the role vertical canopy structure plays in the calculations of LAI and fCover from LiDAR data. One of the most interesting and unexpected results from this study is the high correlation seen between ground LAI and In Canopy High and Low Vegetation Fraction. These correlations suggest more significance can be found when the canopy is segmented by height. While this study used a fairly coarse 3 bin range, changing the number of bins or the calculation used to derive these in canopy values may provide greater insight to LiDAR measurement of forest structure. Three sites in this study

exemplified the type of issues that may arise when vertical canopy structure deviates from the normal vertical distribution of points that is expected from LiDAR data. When canopy point distributions are extremely skewed or have high kurtosis when compared to the expected distribution of LiDAR points, the derived LiDAR values will not correctly estimate values derived from the ground, e.g., a high peak in high vegetation with a lack of middle and lower vegetation, or a high peak in lower vegetation with a lack of middle and high vegetation. Future studies should also consider the inclusion of intermediate returns as a factor in calculating forest metrics. These often overlooked returns may provide greater insight into vertical canopy structure, as well as improve the current measures of LAI and fCover. Lastly, easier, simpler techniques should be developed that allow for quick estimation of tree health with better fit to ground data. Single measures of LAI and fCover from LiDAR did not have great correspondence with ground data, while multivariate measures had varying degrees of success. A simpler technique that is easier to operationalize and implement over a large area is needed. As it stands, the techniques employed in this study can be applied to broader scales, but they are computationally intensive and time consuming.

This study contributes to the growing need for information on the spread of HWA throughout the eastern coast of the United States and helps to characterize the current state of hemlock decline at the southern extent of both the natural range of hemlock and HWA distribution in Georgia. LiDAR and statistical techniques employed here provide a novel way of quantifying the relationship between airborne and ground data, effectively measuring the impact of HWA on hemlock stands. The techniques can be applied on a broad scale to remotely estimate the health of hemlock tree stands. Most importantly, this study provides tools to those involved in research to stop the spread of HWA. Until recently, remote estimation of tree health was

completed using optical methods, but when defoliation of hemlocks starts below the top of the canopy, there is no way to tell from overhead imagery if a stand has been infested. This study demonstrates although the prediction of exact LAI of fCover cannot be achieved with great certainty, when placed into clusters using multivariate analysis, it is possible use LiDAR data to predict up to three health rankings of infestation for tree stands. This analysis may provide a new way to measure infestation, and ultimately gain a perspective on HWA spread and hemlock damage in the eastern United States.

## REFERENCES

- Anderson, R. C. & O. L. Loucks (1979) White-Tail Deer (*Odocoileus Virginianus*) Influence on Structure and Composition of *Tsuga Canadensis* Forests. *Journal of Applied Ecology*, 16, 855-861.
- Applied-Imagery. 2011. Quick Terrain Modeler. Retrieved July, 2011, from <http://www.appliedimagery.com/>.
- Brandtberg, T., T. A. Warner, R. E. Landenberger & J. B. McGraw (2003) Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North America. *Remote Sensing of Environment*, 85, 290-303.
- Brisbin, R. L. (1970) Eastern Hemlock (*Tsuga Canadensis*(L.) Carr.). *USDA Forest Service, American Woods FS-239.*, 8.
- Burns, R. M. & B. H. Honkala, Tech Coords. 1990. Silvics of North America: 1. Conifers. Agricultural Handbook 654., 675 p, 604 - 612. Washinton, D.C.: U.S. Department of Agriculture, Forest Service.
- Carey, J. H. 1993. *Tsuga Canadensis*. In *Fire Effects Information System, [Online]*. Rocky Mountain Research Station, Fire Science Laboratory.: U.S. Department of Agriculture, Forest Service.
- Chen, J. & T. Black (1992) Measuring Leaf Area Index of Plant Canopy with Branch Architecture. *Agricultural and Forest Meteorology*, 57, 12.
- Chen, J. M. (1996) Optically-based methods for measuring seasonal variation of leaf area index in boreal conifer stands. *Agricultural and Forest Meteorology*, 80, 135-163.
- Chen, J. M. & T. A. Black (1991) Measuring leaf area index of plant canopies with branch architecture. *Agricultural and Forest Meteorology*, 57, 1-12.
- Chen, J. M. & J. Cihlar (1996) Retrieving leaf area index of boreal conifer forests using Landsat TM images. *Remote Sensing of Environment*, 55, 153-162.
- CNF, U. 2011. CNF Fact Sheet 2011. Gainsville, GA: USFS.
- Coops, N., T. Hilker, M. Wulder, B. St-Onge, G. Newnham, A. Siggins & J. Trofymow (2007) Estimating canopy structure of Douglas-fir forest stands from discrete-return LiDAR. *Trees - Structure and Function*, 21, 295-310.
- Davis, D. E. 2000. *Where There Are Mountains: An Environmental History of the Southern Appalachians*. Athens, Georgia: University of Georgia Press.

- Dubayah, R., Drake, Jason (2000) Lidar Remote Sensing for Forestry. *Journal of Forestry*, 98, 44-46.
- Hopkinson, C. & L. Chasmer (2009) Testing LiDAR models of fractional cover across multiple forest ecozones. *Remote Sensing of Environment*, 113, 275-288.
- INRA. 2011. Can-Eye Software. Retrieved July, 2011, from <http://www4.paca.inra.fr/can-eye/>.
- IPIX. 2011. IPIX Immersive Imaging. Retrieved July, 2011, from <http://www.ipix.com/>.
- Jolliffe, I. 2005. Principal Component Analysis. In *Encyclopedia of Statistics in Behavioral Science*. John Wiley & Sons, Ltd.
- Jonckheere, I., S. Fleck, K. Nackaerts, B. Muys, P. Coppin, M. Weiss & F. Baret (2004) Review of methods for in situ leaf area index determination: Part I. Theories, sensors and hemispherical photography. *Agricultural and Forest Meteorology*, 121, 19-35.
- Jonckheere, I., K. Nackaerts, B. Muys & P. Coppin (2005) Assessment of automatic gap fraction estimation of forests from digital hemispherical photography. *Agricultural and Forest Meteorology*, 132, 96-114.
- Kane, V. R., J. D. Bakker, R. J. McGaughey, J. A. Lutz, R. F. Gersonde & J. F. Frankling (2010) Examining conifer canopy structural complexity across forest ages and elevations with LiDAR data. *Canadian Journal of Forest Research*, 40, 774-787.
- Lang, A. R. G. & X. Yueqin (1986) Estimation of leaf area index from transmission of direct sunlight in discontinuous canopies. *Agricultural and Forest Meteorology*, 37, 229-243.
- Lefsky, M. A., W. B. Cohen, S. A. Acker, G. G. Parker, T. A. Spies & D. Harding (1999) Lidar Remote Sensing of the Canopy Structure and Biophysical Properties of Douglas-Fir Western Hemlock Forests. *Remote Sensing of Environment*, 70, 339-361.
- Lefsky, M. A., W. B. Cohen, G. G. Parker & D. J. Harding (2002) Lidar Remote Sensing for Ecosystem Studies. *Bioscience*, 52, 19.
- Lim, K., P. Treitz, M. Wulder, B. St-Onge & M. Flood (2003) LiDAR remote sensing of forest structure. *Progress in Physical Geography*, 27, 88-106.
- Lovell, J. L., D. L. B. Jupp, D. S. Culvenor & N. C. Coops (2003) Using airborne and ground-based ranging lidar to measure canopy structure in Australian forests. *Canadian Journal of Remote Sensing*, 29, 607-622.
- Magnussen, S. & P. Boudewyn (1998) Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, 28, 1016-1031.
- McClure, M. S. (1987) Biology and Control of Hemlock Woolly Adelgid. *The Connecticut Agricultural Experiment Station*, 9.

- McClure, M. S. (1989) Evidence of a Polymorphic Life-Cycle in the Hemlock Woolly Adelgid, *Adelges-Tsugae* (Homoptera, Adelgidae). *Annals of the Entomological Society of America*, 82, 50-54.
- McClure, M. S. (1990) Role of Wind, Birds, Deer, and Humans in the Dispersal of Hemlock Woolly Adelgid (Homoptera, Adelgidae). *Environmental Entomology*, 19, 36-43.
- Means, J. E., S. A. Acker, D. J. Harding, J. B. Blair, M. A. Lefsky, W. B. Cohen, M. E. Harmon & W. A. McKee (1999) Use of Large-Footprint Scanning Airborne Lidar To Estimate Forest Stand Characteristics in the Western Cascades of Oregon. *Remote Sensing of Environment*, 67, 298-308.
- Miller, J. (1967) A Formula for Average Foliage Density. *Australian Journal of Botany*, 15, 4.
- Morsdorf, F., E. Meier, B. Kötz, K. I. Itten, M. Dobbertin & B. Allgöwer (2004) LIDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. *Remote Sensing of Environment*, 92, 353-362.
- Næsset, E. (1997) Determination of mean tree height of forest stands using airborne laser scanner data. *Isprs Journal of Photogrammetry and Remote Sensing*, 52, 49-56.
- Nelson, R., W. Krabill & G. Maclean (1984) Determining Forest Canopy Characteristics Using Airborne Laser Data. *Remote Sensing of Environment*, 15, 201-212.
- Nelson, R., W. Krabill & J. Tonelli (1988) Estimating Forest Biomass and Volume Using Airborne Laser Data. *Remote Sensing of Environment*, 24, 247-267.
- Orwig, D. A. & D. R. Foster (1998) Forest response to the introduced hemlock woolly adelgid in southern New England, USA. *Journal of the Torrey Botanical Society*, 125, 60-73.
- Orwig, D. A., D. R. Foster & D. L. Mausel (2002) Landscape patterns of hemlock decline in New England due to the introduced hemlock woolly adelgid. *Journal of Biogeography*, 29, 1475-1487.
- Popescu, S. C., R. H. Wynne & R. F. Nelson (2003) Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing*, 29, 564-577.
- QCoherent. 2011. LP360. Retrieved July, 2011, from <http://www.qcoherent.com/>.
- Riaño, D., F. Valladares, S. Condés & E. Chuvieco (2004) Estimation of leaf area index and covered ground from airborne laser scanner (Lidar) in two contrasting forests. *Agricultural and Forest Meteorology*, 124, 269-275.
- Roberts, S. W., R. Tankersley Jr & K. H. Orvis (2009) Assessing the Potential Impacts to Riparian Ecosystems Resulting from Hemlock Mortality in Great Smoky Mountains National Park. *Environmental Management*, 44, 335-345.

- Stadler, B., T. Müller, D. Orwig & R. Cobb (2005) Hemlock Woolly Adelgid in New England Forests: Canopy Impacts Transforming Ecosystem Processes and Landscapes. *Ecosystems*, 8, 233-247.
- Trimble. 2011a. Nomad Handheld Field Computer. Retrieved July, 2011, from <http://www.trimble.com/Outdoor-Rugged-Computers/nomad.aspx>.
- . 2011b. Pathfinder ProXH. Retrieved July, 2011, from <http://www.trimble.com/pathfinderproxh.shtml>.
- USFS. 2011. Hemlock Woolly Adelgid., ed. U. F. Service. US Forest Service.
- Watson, D. J. (1947) Comparative physiological studies in the growth of field crops. I. Variation in net assimilation rate and leaf area between species and varieties, and within and between years. *Annals of Botany*, 11.
- Wehr, A. & U. Lohr (1999) Airborne laser scanning - an introduction and overview. *Isprs Journal of Photogrammetry and Remote Sensing*, 54, 68-82.
- Weiss, M. 2010. *Can-Eye Users Guide*. Domaine Saint-Paul: French National Institute of Agronomical Research.
- Weiss, M., F. Baret, G. J. Smith, I. Jonckheere & P. Coppin (2004) Review of methods for in situ leaf area index (LAI) determination: Part II. Estimation of LAI, errors and sampling. *Agricultural and Forest Meteorology*, 121, 37-53.
- Wold, S., K. Esbensen & P. Geladi (1987) Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2, 37-52.
- Yamasaki, M. D., Richard M.; Lanier, John W. 2000. Wildlife habitat associations in eastern hemlock - birds, smaller mammals, and forest carnivores. In *Proceedings: Symposium on sustainable management of hemlock ecosystems in eastern North America. Gen. Tech. Rep. NE-267.*, ed. K. A. S. McManus, Kathleen S.; Souto, Dennis R. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station. 135-143.
- Zwick, W. R. & W. F. Velicer (1986) Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432-442.